

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-339346

(43)Date of publication of application : 08.12.2000

(51)Int.Cl.

G06F 17/30

(21)Application number : 2000-032625

(71)Applicant : HITACHI LTD

(22)Date of filing : 03.02.2000

(72)Inventor : INABA YASUHIKO
 MATSUBAYASHI TADATAKA
 TADA KATSUMI
 OKAMOTO TAKUYA
 SUGAYA NATSUKO
 GOCHI YOSUKE

(30)Priority

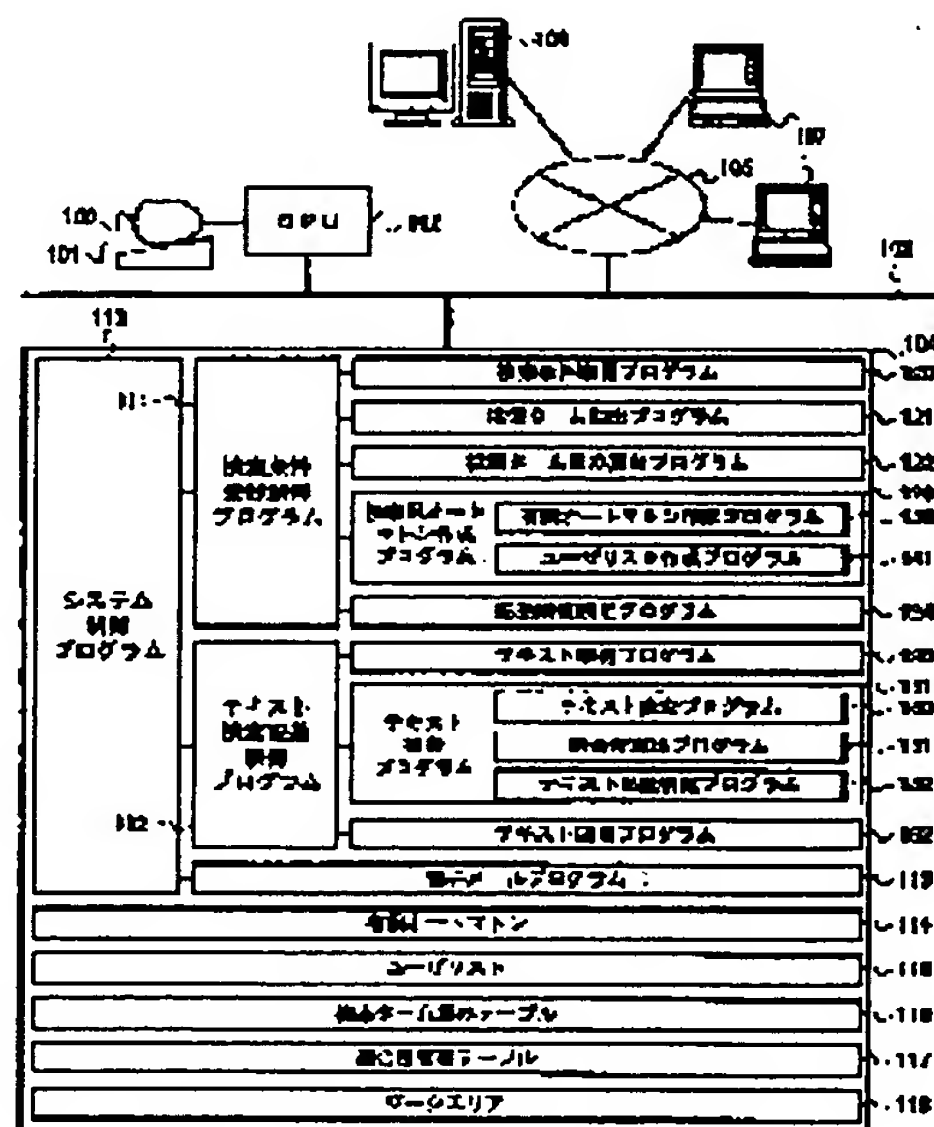
Priority number : 11075005 Priority date : 19.03.1999 Priority country : JP

(54) METHOD AND DEVICE FOR RETRIEVING AND DELIVERING SIMILAR DOCUMENT, AND STORAGE MEDIUM STORED WITH SIMILAR DOCUMENT RETRIEVAL DELIVERY PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a system capable of retrieving information desired by a user without needing the preparation of a complicated retrieval conditional expression and also fast delivering only a text including information desired by the user regardless of the number of users.

SOLUTION: This device has a retrieval condition registration step which generates a finite automaton 114 collating a retrieval term extracted from a registered retrieval condition by a user 107 and a text retrieval delivery step which calculates the goodness of fit to respective retrieval conditions designated by plural users in such a manner that the automaton 114 retrieves a text distributed from a news distribution source 106 and distributes the text to a user who designates a retrieval condition satisfying a prescribed condition.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of

rejection]

[Date of extinction of right]

Copyright (C): 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号
特開2000-339346
(P2000-339346A)

(43)公開日 平成12年12月 8 日 (2000.12.8)

(51)Int.Cl.⁷

識別記号

F I

テーマコード(参考)

G 0 6 F 17/30

G 0 6 F 15/403
15/40
15/403

3 5 0 C 5 B 0 7 5
3 7 0 A
3 3 0 C

審査請求 未請求 請求項の数24 O L (全 56 頁)

(21)出願番号 特願2000-32625(P2000-32625)

(22)出願日 平成12年 2 月 3 日 (2000. 2. 3)

(31)優先権主張番号 特願平11-75005

(32)優先日 平成11年 3 月19日 (1999. 3. 19)

(33)優先権主張国 日本 (J P)

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目 6 番地

(72)発明者 稲場 靖彦

神奈川県川崎市幸区鹿島田890番地 株式

会社日立製作所システム開発本部内

(72)発明者 松林 忠孝

神奈川県川崎市幸区鹿島田890番地 株式

会社日立製作所システム開発本部内

(74)代理人 100075096

弁理士 作田 康夫

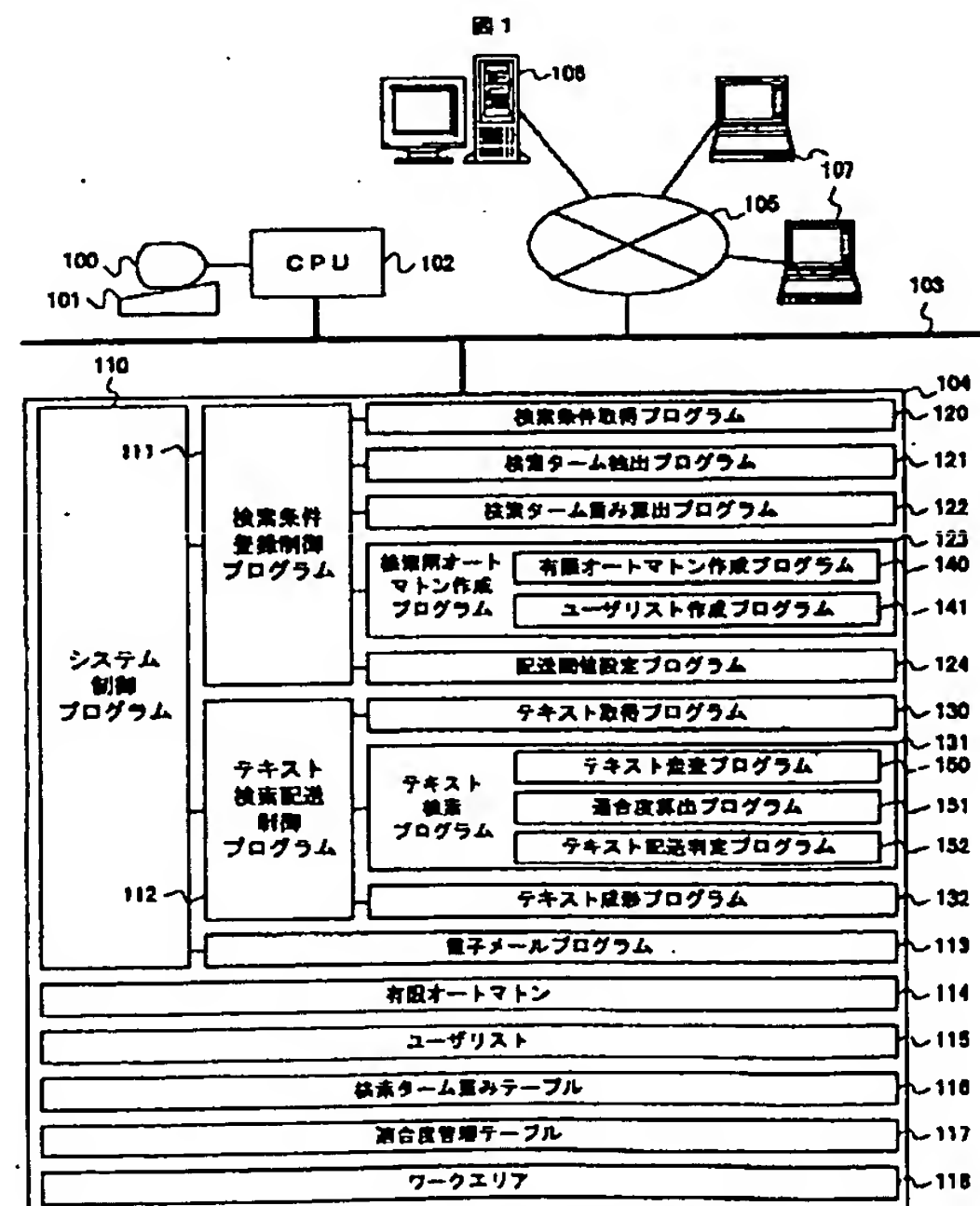
最終頁に続く

(54)【発明の名称】 類似文書検索配送方法および装置並びに類似文書検索配送プログラムを格納した記憶媒体

(57)【要約】

【課題】本発明の課題は、複雑な検索条件式の作成を必要とせずにユーザが所望する情報を検索できるようにするとともに、ユーザが真に所望する情報を含むテキストだけをユーザ数によらず高速に配送することができるシステムを提供することである。

【解決手段】ユーザ107により登録された検索条件から抽出された検索タームを照合する有限オートマトン114を生成する検索条件登録ステップと、ニュース配信元106から配信されたテキストを該有限オートマトン114で検索することにより、複数のユーザに指定された検索条件それぞれに対する適合度を算出し、該適合度が所定の条件を満足する検索条件を指定したユーザに対して該テキストを配信するテキスト検索配送ステップを有する。



(本発明の第一の実施例の構成を示す図)

【特許請求の範囲】

【請求項 1】複数のユーザから入力した検索条件を登録する検索条件登録ステップと、入力した文書情報のテキストを対象として、前記検索条件を満足するテキストを検索し、該ユーザに配送する検索配送ステップとを有する文書検索配送方法は、さらに以下を含む、
前記検索配送ステップでは、前記検索条件に対する適合度を上記テキストから算出し、該適合度により該検索条件の成否を判定し、成立した該検索条件に対応する前記ユーザに対して、該テキストを配送する。

【請求項 2】請求項 1 記載の文書検索配送方法において、
上記検索条件登録ステップは、
各ユーザが登録した検索条件中に記述された単語や文章あるいは文書（以下、種文書と呼ぶ）を読み込む種文書読み込みステップと、
上記種文書読み込みステップで登録された種文書を解析し、検索に使用する検索タームを抽出する検索ターム抽出ステップと、
上記検索ターム抽出ステップで抽出された各検索タームに対して、上記種文書読み込みステップで読み込まれた種文書を指定したユーザのユーザ識別子を登録するユーザ識別子登録ステップと、
各ユーザが登録した検索条件中に記述された配送条件を登録する配送条件登録ステップを有し、
前記テキスト検索配送ステップは、
該テキストに対して、前記検索ターム抽出ステップで抽出された検索タームを検索する検索ターム検索ステップと、
上記検索ターム検索ステップで検索された該検索ターム毎に登録された前記ユーザ識別子を取得するユーザ識別子取得ステップと、
上記検索ターム検索ステップで検索された該検索タームの出現情報と、上記ユーザ識別子取得ステップで取得されたユーザ識別子を用いて、所定の算出式に基づいて、該テキストの各検索条件に対する適合度を算出する適合度算出ステップと、
上記適合度算出ステップで算出された適合度が前記配送条件登録ステップで登録された配送条件を満たすテキストを、該当するユーザに配送するテキスト配送ステップを有することを特徴とする文書検索配送方法。
【請求項 3】請求項 2 記載の文書検索配送方法における検索条件登録ステップにおいて、
前記検索ターム抽出ステップで抽出された該検索タームに対する重みを算出する検索ターム重み算出ステップを有し、
前記テキスト検索配送ステップにおいて、
前記検索ターム検索ステップの代わりに、前記検索ターム抽出ステップで抽出された検索タームの該テキストにおける出現回数を計数する出現回数計数ステップと、
前記適合度算出ステップの代わりに、上記検索ターム重

み算出ステップで算出された検索ターム重みと上記出現回数計数ステップで計数された該検索ターム出現回数を用いることにより、各検索条件に対する該テキストの適合度を算出する適合度算出ステップを有することを特徴とした文書検索配送方法。

【請求項 4】請求項 2 および 3 記載の文書検索配送方法におけるテキスト検索配送ステップにおいて、
検索条件に対する取得したテキストの適合度を算出する順番を制御する適合度算出順序制御ステップを有することを特徴とする文書検索配送方法。

【請求項 5】請求項 4 記載の文書検索配送方法における適合度算出順序制御ステップにおいて、
予め定められた配信優先度、ユーザが指定した配信優先度あるいは配信サービスの配信優先度を含む契約料金に従って前記適合度を算出する順番を決定する適合度算出順序決定ステップを有することを特徴とする文書検索配送方法。

【請求項 6】請求項 3 記載の文書検索配送方法における検索ターム重み算出ステップにおいて、
前記検索ターム抽出ステップで抽出された該検索タームが出現するテキスト数を計数する検索ターム出現テキスト数計数ステップと、
上記検索ターム出現テキスト数計数ステップで計数されたテキスト数を用いて該検索タームの重みを算出する検索ターム出現テキスト数利用重み算出ステップを有することを特徴とする文書検索配送方法。

【請求項 7】請求項 1 記載の文書検索配送方法において、請求項 1 記載のテキスト検索配送ステップで行う処理に加え、取得した全テキスト、および全テキストに対する各ユーザの検索条件との適合度を保存する、テキスト検索配送ステップを有することを特徴とする文書検索配送方法。

【請求項 8】請求項 7 記載の文書検索配送方法において、前記テキスト検索配送ステップにおいて保存した内容をユーザに提示する、配送閾値設定支援情報提示ステップを有することを特徴とする文書検索配送方法

【請求項 9】請求項 8 記載の文書検索配送方法において、前記配送閾値設定支援情報提示ステップにおいてユーザに提示する内容として、ユーザに配送したテキストの、適合度、配送時刻、配送時点での設定配送閾値を該ユーザに一覧提示する、配送閾値設定支援情報提示ステップを有することを特徴とする文書検索配送方法

【請求項 10】請求項 8 記載の文書検索配送方法において、前記配送閾値設定支援情報提示ステップにおいてユーザに提示する内容として、過去に算出した適合度別のテキスト数を該ユーザに提示する、配送閾値設定支援情報提示ステップを有することを特徴とする文書検索配送方法

【請求項 11】請求項 8 記載の文書検索配送方法において、前記配送閾値設定支援情報提示ステップにおいてユ

ーザに提示する内容として、過去に取得したテキストの適合度を、受信時刻に対してプロットしたグラフの形で該ユーザに提示する、配送閾値設定支援情報提示ステップを有することを特徴とする文書検索配送方法。

【請求項 12】請求項 1 記載の文書検索配送方法において、予め設定された、テキストを配送するか否かを決定する基準の値である配送閾値と、テキストの適合度算出結果とを比較して、テキストの適合度が上回る場合に、テキストを該ユーザに配送するテキスト配送ステップを有することを特徴とする文書検索配送方法。

【請求項 13】請求項 12 記載の文書検索配送方法において、配送閾値をユーザ毎に設定する、配送閾値設定ステップを有することを特徴とする文書検索配送方法。

【請求項 14】請求項 13 記載の文書検索配送方法において、請求項 7 記載のテキスト検索配送ステップを有し、前記テキスト検索配送ステップにおいて保存したテキストを走査し、各ユーザの設定した検索条件との適合度を算出することにより、配送閾値を各ユーザ毎に算出し設定する、配送閾値自動設定ステップを有することを特徴とする文書検索配送方法。

【請求項 15】請求項 13 記載の文書検索配送方法において、請求項 9、または請求項 10、または請求項 11 記載の配送閾値設定支援情報提示ステップを有し、請求項 9、または請求項 10、または請求項 11 における提示情報を、ユーザが配送閾値を設定する際に提示する、配送閾値設定支援情報提示ステップを有することを特徴とする文書検索配送方法。

【請求項 16】請求項 13 記載の文書検索配送方法において、請求項 7 記載のテキスト検索配送ステップを有し、前記テキスト検索配送ステップにおいて保存した情報をもとに、過去の一定期間内の全テキストのうち、請求項 1 記載の適合度が新しく設定した配送閾値を超えるテキストをユーザに提示する、配送閾値設定試行ステップを有することを特徴とする文書検索配送方法。

【請求項 17】請求項 12 または請求項 13 記載の文書検索配送方法において、請求項 7 記載のテキスト検索配送ステップを有し、前記テキスト検索配送ステップにおいて保存した情報をもとに、適合度が請求項 12 記載の配送閾値を超えなかったテキストを対象に、ユーザが希望する配送数、または配送頻度を満たす数まで、適合度の上位のテキストを追加配送する、テキスト追加配送ステップを有することを特徴とする文書検索配送方法。

【請求項 18】請求項 13 記載の文書検索配送方法において、請求項 7 記載のテキスト検索配送ステップを有し、前記テキスト検索配送ステップにおいて保存した情報をもとに、ユーザに配送閾値の修正を促すメッセージを提示する、配送閾値修正提案提示ステップを有することを特徴とする文書検索配送方法。

【請求項 19】請求項 1 記載の文書検索配送方法におい

て、一定期間内あるいはユーザが設定する期間内に該ユーザが希望する配送テキスト数を設定する、希望配送数設定ステップを有することを特徴とする文書検索配送方法。

【請求項 20】請求項 19 記載の文書検索配送方法において、請求項 7 記載のテキスト検索配送ステップを有し、前記希望配送数設定ステップにおいて設定されたテキスト数と、前記テキスト検索配送ステップにおいて保存した情報をもとに、所定の時間毎にユーザ毎の配送閾値を設定する、配送閾値自動設定ステップを有することを特徴とする文書検索配送方法。

【請求項 21】請求項 1 記載の文書検索配送方法において、請求項 13 記載の配送閾値設定ステップと、請求項 19 記載の希望配送数設定ステップを有し、前記配送閾値設定ステップと、前記希望配送数設定ステップのどちらを使用するかをユーザが選択する、配送条件設定モード選択ステップを有することを特徴とする文書検索配送方法。

【請求項 22】複数のユーザから入力した検索条件を登録する検索条件登録手段と、入力した文書情報のテキストデータを対象として、前記検索条件を満足するテキストを検索し、該ユーザに配送する検索配送手段とを有する文書検索配送装置は、さらに以下を含む、前記検索配送手段では、前記検索条件に対する適合度を上記テキストから算出し、該適合度により該検索条件の成否を判定し、成立した該検索条件に対応するユーザに対して、該テキストを配送する。

【請求項 23】複数のユーザから入力した検索条件を登録する検索条件登録ステップと、入力した文書情報のテキストデータを対象として、前記検索条件を満足するテキストを検索し、該ユーザに配送する検索配送ステップとを有する文書検索配送プログラムは、さらに以下を含む、前記検索配送ステップでは、前記検索条件に対する適合度を上記テキストから算出し、該適合度により該検索条件の成否を判定し、成立した該検索条件に対応するユーザに対して、該テキストを配送する。

【請求項 24】複数のユーザから入力した検索条件を登録する検索条件登録ステップと、入力した文書情報のテキストデータを対象として、前記検索条件を満足するテキストを検索し、該ユーザに配送する検索配送ステップとを有する文書検索配送プログラムを格納した計算機読み取り可能な記憶媒体、前記文書検索配送プログラムは、さらに以下を含む、前記検索配送ステップでは、前記検索条件に対する適合度を上記テキストから算出し、該適合度により該検索条件の成否を判定し、成立した該検索条件に対応するユーザに対して、該テキストを配送する。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、電子化された文書を、ユーザが予め登録しておいた検索条件で検索し、条件を満たす文書をそのユーザに配送する文

【0002】書検索配送技術に関する。

【従来技術】 近年、電子メールや電子ニュース等により大量の電子化文書（以下、テキストと呼ぶ）が時々刻々ユーザへ配信されるようになってきた。また、WWW（World Wide Web）を利用して情報発信を行なう情報源が急増しており、これらの情報源から情報収集ロボット等を用いて収集されるテキストも膨大な量となっている。このため、これらのテキストの中から、ユーザが求める情報を含むテキストを検索し、これを該ユーザに配信する文書検索配送システムへのニーズが高まっている。

【0003】この文書検索配送システムとして、「特開平10-27182」（以下、従来技術1と呼ぶ）が開示されている。従来技術1では、複数ユーザの検索条件式を統合し、1度のテキスト走査により複数ユーザの条件式を処理している。

【0004】しかし、従来技術1では、ユーザが検索条件式を作成する必要があるため、次の二つの問題がある。

【0005】まず、第一の問題は、あまり使用されない希少な単語を検索条件として指定した場合や、一般的な単語であっても複雑に組み合わせた検索条件を指定した場合には、検索漏れが発生してしまうことである。

【0006】そして、第二の問題は、上記第一の問題とは逆に、一般的な単語だけを指定した単純な検索条件式を指定した場合には、検索の目的には適合しない文書

（以下、検索ノイズと呼ぶ）を検索結果に多く含んでしまうことである。このため目的の文書を容易に入手することができないという問題がある。

【0007】すなわち、検索漏れがなく、検索ノイズの少ない検索結果を取得するために、適切な検索条件式の作成が困難であるという問題がある。

【0008】テキストデータベースに登録されている文書（以下、登録文書と呼ぶ）の中から、所望の情報を含む文書を検索する文書検索システムにおいて、上記二つの問題を改善する技術として、「特願平10-148721」（以下、従来技術2と呼ぶ）が開示されている。

【0009】従来技術2では、検索条件として例示されたテキスト（以下、種文書と呼ぶ）からキーワード（従来技術2では特徴文字列と呼んでいる）を抽出し、登録文書の該種文書に対する適合度を算出する。

【0010】従来技術2を用いることにより、ユーザは所望の情報を含む種文書を例示するだけで、どの検索チームを使用して検索条件式を作成すればよいのか悩むことなく検索の実行を指示し、適合度によりソートされた検索結果上位から順に閲覧することで、検索結果に検索ノイズが多く含まれていても所望の情報を容易に入手す

ることができるようになる。

【0011】

【発明が解決しようとする課題】 以下、上記二つの従来技術について、その概要を説明し、その問題点を述べる。

【0012】まず従来技術1の概要について、図2を用いて説明する。

【0013】本図は、文書検索配送システムに3人のユーザの検索条件式、ユーザ1：「“文書”と“検索”が含まれる文書」、ユーザ2：「“登山”が含まれる文書」およびユーザ3：「“検索”と“登録”が含まれる文書」が登録されており、収集したテキスト：「文書を検索する」を走査し、これら3条件に合致するか否かを判断する例である。

【0014】予め、各ユーザが登録した検索条件式を解析し、検索チーム“文書”、“検索”、“登山”、“登録”を抽出する。

【0015】そして、抽出された検索チームの数を、各ユーザ毎に検索チーム数カウントテーブルに格納する。例えば、ユーザ1が登録した検索条件式ユーザ1：

「“文書”と“検索”が含まれる文書」からは、“文書”と“検索”の2つの検索チームが抽出されるため、検索チーム数カウントテーブルの対応する箇所に“2”が格納される。同様に、ユーザ2、ユーザ3に対応する箇所に“1”、“2”がそれぞれ格納される。

【0016】次に、抽出された全ての検索チームを照合する有限オートマトンが生成される。

【0017】図2における有限オートマトンでは、円形は有限オートマトンの状態を、矢印は状態遷移を表している。各矢印に付記された文字はこれに対応した状態遷移が起きる入力文字を、各円形の内部に記された数値は同状態の状態番号を示す。なお本図では、有限オートマトンに示されていない文字が入力された場合（以下、フェイルと呼ぶ）の初期状態への遷移を表す矢印は省略している。

【0018】そして、それぞれの検索チームを指定したユーザの識別子をユーザリストとして、有限オートマトンの対応する検索チーム検出状態に接続する。本図では、例えば“検索”を照合するとその末尾の状態4からユーザリストが参照され、“検索”を指定したユーザが“ユーザ1”および“ユーザ3”であることが検出される。

【0019】次に、テキスト：「文書を検索する」を図2に示した有限オートマトンを用いて走査する処理について説明する。本例では、“文書”および“検索”と一致する部分文字列がテキスト中に出現していることが検出される。本図に示した有限オートマトンで、末尾の状態に“○”が付記されている検索チームはテキスト中に一致する部分文字列が出現したことを示す。本例では、“文書”および“検索”と一致する部分文字列がテキス

ト中に出現しているので、その末尾の状態である状態 2 と状態 4 に“○”が付記される。

【0020】そして、テキスト中の部分文字列と一致した検索タームの出現数をユーザ毎にカウントし、検索ターム出現数カウントテーブルに格納する。例えば、ユーザ 1 に対しては、“文書”と“検索”が一致しているので“2”とカウントし、ユーザ 3 に対しては、“検索”だけが一致しているので“1”とカウントする。ユーザ 2 はテキスト中の検索タームと一致する部分文字列が現れなかったのでカウントを行わず“0”のままである。

【0021】最後に、検索条件式から抽出された検索ターム数をユーザ別に記憶した検索ターム数カウントテーブルと、テキスト中の部分文字列として出現している検索ターム数を記憶した検索ターム出現数カウントテーブルとを比較し、一致している場合には、ユーザが指定した検索条件式が合致したものとみなして該テキストを該ユーザへ配送する。図 2 でユーザ 1 の場合は検索ターム数が 2 で一致しているためテキストを配送するが、ユーザ 2 およびユーザ 3 は検索ターム数が一致していないので配送しない。

【0022】以上が従来技術 1 の概要である。

【0023】本技術により、与えられた検索条件式に適合するテキストを 1 度の走査によりユーザに配送する文書検索配送システムを実現することができる。

【0024】しかし、従来技術 1 では、ユーザが検索条件式を作成する必要があるため、適切な検索条件式の作成が困難であるという問題がある。

【0025】この問題を改善する技術として、文書検索システムにおいては、前述の従来技術 2 が提案されている。

【0026】以下、従来技術 2 の概要について、図 20 を用いて説明する。

【0027】図 20 は、種文書「・・・。携帯電話の使用時のマナーが問題になる。・・・」からキーワード

(以下、従来技術 2 での呼称“特徴文字列”に合わせて記述する)を抽出する場合の例である。

【0028】まず、ステップ 1910 で単一文字種文字列抽出プログラムを起動し、種文書 1920 を漢字やカタカナなどの文字種の境界で分割することにより、単一の文字種で構成される文字列(以下、単一文字種文字列と呼ぶ)1921 を抽出する。

【0029】次に、各単一文字種文字列の文字種を判定(ステップ 1911)し、複合語を構成する可能性の高い漢字文字列とカタカナ文字列に対して、分割確率比較特徴文字列抽出プログラムを実行することで複合語の分割処理を行い、特徴文字列を抽出する。また、複合語を構成する可能性の低いその他の文字種の文字列については、単一文字種文字列そのものを特徴文字列として抽出する(ステップ 1912)。

【0030】そしてステップ 1913 にて、抽出された

特徴文字列をワークエリアに格納する。

【0031】以上が、従来技術 2 における種文書から特徴文字列を抽出する処理の例である。

【0032】従来技術 2 では、本処理により抽出された特徴文字列の種文書内出現回数とテキストデータベース内の各文書における出現回数から各文書の適合度を算出し、適合度の降順に表示する。この適合度の算出方法としては、従来技術 2 で説明されているので、ここでは説明を省略する。

【0033】以上が従来技術 2 の概要である。

【0034】従来技術 2 では、種文書中から特徴文字列(キーワード)を抽出し、このキーワードを用いて種文書に対する登録文書の適合度を算出する。したがって、ユーザは所望の情報を含む文書を例示し、適合度の降順に検索結果を参照することで、テキストデータベースから所望の情報を含むテキストを入手することができるようになる。

【0035】したがって、本従来技術 2 を従来技術 1 に適用することにより、種文書中から抽出したキーワードを一つでも含む登録文書に対して、種文書に対する適合度を付与することで、従来技術 1 の第一の問題を改善することができる。

【0036】しかし、従来技術 1 では時々刻々と配信されるテキスト 1 件毎に検索条件に対する適否を判定し、適合したテキストを即時にユーザへ配送する。このため、従来技術 2 を用いて検索条件に対する各テキストの適合度を算出することができても、他のテキストとの適合度と比較することができない。

【0037】すなわち、従来技術 2 を単純に従来技術 1 の文書検索配送システムに適用したとしても、検索結果を適合度の降順にソートして配送することができないため、従来技術 1 の第二の問題が改善されないことになる。

【0038】また、従来技術 2 ではキーワードの抽出や適合度の算出にテキストデータベース中の統計情報を利用することが可能であるが、従来技術 1 の文書検索配送システムでは配信されたテキストを貯えないため、統計情報の取得が困難であるという問題が発生する。

【0039】本発明の目的は、複数のユーザが入力した検索条件に基づいて入力したテキストに対し算出し、前記検索条件が成立するかを判定し、成立する検索条件に対応するユーザに該入力テキストを配送することにより、ユーザが所望とするテキストを配送することができるシステムを提供することにある。

【0040】

【課題を解決するための手段】上記課題を改善するため、本発明に示す文書検索配送方法では、以下に示すステップからなる処理により、ユーザが所望する情報を含む文書(以下、種文書と呼ぶ)に対する取得されたテキストの適合度を判別し、ユーザへ配布する。

【0041】すなわち、本発明による文書検索配送方法では、複数のユーザから入力した検索条件を登録する検索条件登録ステップと、文書情報のテキストデータを対象として、前記検索条件を満足するテキストを検索し、該ユーザに配送する検索配送ステップとを有する。

【0042】さらに、前記検索配送ステップでは、前記検索条件に対する該テキストの合致する割合（以下、適合度と呼ぶ）を上記テキストから算出し、該適合度により該検索条件の成否を判別し、成立した該検索条件に対応する前記ユーザに対して、該テキストを配送する。

【0043】このようにすることにより、ユーザが所望する情報を含む文書に対する取得されたテキストの適合度を判別し、ユーザへ配布することが可能となる。

【0044】上記ユーザから入力した検索条件は、そのユーザが所望する文書で与えても良い。その場合、本発明では、その文書に基づいた検索条件を生成する。

【0045】

【発明の実施の形態】 以下、本発明の第一の実施例について図を用いて説明する。

【0046】最初に、第一の実施例の概略を図8および図9を用いて説明する。

【0047】まず、検索条件登録処理の概要を図8を用いて説明する。本図は、ユーザ1：“文書検索システムとしては、HiRetrievalがよく知られて・・・”、ユーザ2：“冬山登山における防寒対策としては、・・・”およびユーザ3：“システムの文書登録性能を評価すると、・・・”という3人のユーザの検索条件が登録される場合の例である。

【0048】まず、各ユーザにより登録された検索条件から検索タームを抽出する。本例では、ユーザ1が登録した検索条件ユーザ1：“文書検索システムとしては、HiRetrievalがよく知られて・・・”からは、“文書”、“検索”、“システム”および“HiRetrieval”の4つの検索タームが抽出される。同様に、ユーザ2が登録した検索条件からは、“冬山”、“登山”、“防寒”および“対策”の4つの検索タームが抽出され、ユーザ3が登録した検索条件からは、“システム”、“文書”、“登録”、“性能”および“評価”の5つの検索タームが抽出される。

【0049】そして、抽出された全ての検索タームに対して所定の算出式を用いて重みを算出し、検索ターム重みテーブルに格納する。

【0050】重みの算出には、例えば、“Information Retrieval”, William B. Frakes /Ricardo Baeza-Yates, Prentice Hall PTR, 1992, p.p. 363 ~ 391, “Ranking Algorithm”」（以下、従来技術3と呼ぶ）に開示されているIDF(Inverted Document Frequency)算出式

(1)を用いてもよい。

【0051】

【数1】

$$IDF(i) = 1 + \log_2 \frac{N}{n(i)} \quad \dots (1)$$

(IDF算出式の例)

【0052】ここで、IDF(i)は検索ターム(i)のIDFを示し、Nはテキストデータベース中の全文書数を示す。また、n(i)は検索ターム(i)の出現文書数を示す。このIDF算出式(1)を用いると、例えば、10万件のテキストが格納されているテキストデータベースにおいて、ユーザ1の登録した検索条件から抽出された検索ターム“文書”が2,000件のテキストに出現している場合には、検索ターム“文書”の重みは“6.6”と算出され、検索ターム重みテーブルに格納される。

【0053】ここで、テキストデータベースに登録されている全文書数および検索タームの出現している文書数は、それぞれニュース配信元から配信されたテキスト数Nおよび該検索ターム(i)が出現したテキスト数n(i)を計数しておき、これを利用してもよい。あるいは、配信されたテキスト、あるいは異なるテキストが登録されているテキストデータベースを参照することにより求めてもよい。これにより、文書検索配送システムに対して、統計情報を用いた適合度算出処理を取得することが可能となる。

【0054】次に、抽出された全ての検索タームを照合する有限オートマトンが作成される。図8における有限オートマトンでは、円形は有限オートマトンの状態を、矢印は状態遷移を表している。各矢印に付記された文字はこれに対応した状態遷移が起きる入力文字を、各円形の内部に記された数値は同状態の状態番号を示す。なお本図では、有限オートマトンに示されていない文字が入力された場合（以下、フェイルと呼ぶ）の初期状態への遷移を表す矢印は省略している。また本図に示した有限オートマトンは、本例で作成される有限オートマトンの一部を省略している。

【0055】そして、それぞれの検索タームが抽出された検索条件を指定したユーザの識別子をユーザリストとして、有限オートマトンの対応する検索ターム検出状態に接続する。本図では、例えば“検索”を照合するとその末尾の状態4からユーザリストが参照され、“検索”を指定したユーザが“ユーザ1”であることが検出される。

【0056】また、各ユーザにより登録された検索条件から配送閾値を抽出する。本例では、ユーザ1が登録した検索条件からは配送閾値8.0が抽出される。同様に、ユーザ2およびユーザ3が登録した検索条件からは、それぞれ3.5、7.0が抽出される。

【0057】そして、抽出された配送閾値はそれぞれユーザ識別子に対応づけて、適合度管理テーブルに格納される。

【0058】次に、テキスト検索処理の概要を図9を用

いて説明する。本図は、テキスト：“膨大な文書の中から目的の文書を検索する。”を図8に示した有限オートマトンを用いて走査し、適合する検索条件を検索する場合の例である。

【0059】まず、図9に示した有限オートマトンを用いて各検索タームのテキスト中での出現回数を計数する。本図に示した有限オートマトンでは、末尾の状態に各検索タームのテキスト中での出現回数が付記されている。本例では、検索ターム“文書”に一致する部分文字列がテキスト中に2回出現しているので、その末尾の状態である状態2に“2”が付記され、検索ターム“検索”に一致する部分文字列がテキスト中に1回出現しているので、その末尾の状態である状態4に“1”が付記される。

【0060】そして、テキスト中の部分文字列と一致した全ての検索タームに対して、その末尾の状態に接続されたユーザリストを参照し、ユーザの登録した各検索条件に対するテキストの適合度を算出する。ここで各検索条件に対するテキストの適合度算出式としては、従来技術3に開示されている適合度算出式(2)を用いてもよい。

【0061】

【数2】

$$Similarity = \sum_i \left((C + IDF(i)) \times \left(K + (1-K) \frac{freq(i)}{\max(freq(j))} \right) \right) \dots (2)$$

(適合度算出式の例)

【0062】ここで、Qはユーザが登録した検索条件から抽出された検索ターム数を示し、CおよびKは定数を示す。また、IDF(i)は検索ターム(i)のIDFを示し、freq(i)は該テキストにおける検索ターム(i)の出現回数を示す。また、max(freq(j))は、該テキストにおける検索タームの出現回数のうち、最大の値を示す。

【0063】本図に示した例では、定数CおよびKの値を0とすると、ユーザ1の登録した検索条件に対するテキストの適合度は“9.2”(6.6×2/2+5.1×1/2=9.15)のように算出される。同様に、ユーザ2の登録した検索条件に対する適合度は“0”、ユーザ3の登録した検索条件に対する適合度は“6.6”と算出される。

【0064】そして、算出された適合度が各ユーザにより予め設定された配送閾値を超えている場合には、該当する検索条件を登録したユーザへテキストを配送する。本図に示した例の場合、ユーザ1の配送閾値“8.0”に対してテキストの適合度は“9.2”であるため、ユーザ1に該テキストが配送される。しかし、ユーザ2、ユーザ3が設定した配送閾値“3.5”、“7.0”に対して、テキストの適合度はそれぞれ“0.0”、“6.6”であり、いずれも配送閾値も超えていない。したがって、該テキストはユーザ2、ユーザ3には配送されない。

【0065】以上のように、本実施例では、有限オート

マトンを用いてテキストを走査し、テキスト中に現れた検索タームの出現回数を計数する。そして、ユーザリストを参照しながら各ユーザの登録した各検索条件に対する該テキストの適合度を算出し、適合度がユーザの設定した配送条件を満たしている検索条件を指定したユーザに対して、該テキストを配送する。

【0066】この結果、テキストのただ一度の走査で複数ユーザの検索条件に対する適合度を算出することが可能であり、またユーザ毎に配送閾値との比較を行うため、多くの情報を要求するユーザに対しては適合度の低いテキストまでも配送することが可能となり、重要な情報だけを要求するユーザに対しては適合度の高いテキストを配送することが可能となる。

【0067】以下、本発明の第一の実施例について図1を用いて詳細に説明する。

【0068】本発明を適用した文書検索配送システムの第一例は、ディスプレイ100、キーボード101、中央演算処理装置(CPU)102、主メモリ104およびこれらを結ぶバス103から構成される。

【0069】また、バス103には、LAN(Local Area Network)等の通信回線105を介して、ニュースを配信するニュース配信元106や文書検索配送システムを利用するユーザ107が接続されている。ニュース配信元106は電子メールや電子ニュース等を用いてニュースデータを電子化したテキストを本システムへ配信したり、インターネットを介してテキストを提示する。ユーザ107は電子メールを用いて検索条件を本システムへ登録する。本システムからは上記検索条件に基づいて検索された上記テキストが電子メールを用いて該当ユーザへ配送される。

【0070】以下、本実施例では、ニュース配信元106は電子メール等を用いて本システムにテキストを配信するものとして述べるが、ニュース配信元106はインターネット上に提示するだけで、テキストの収集は情報収集ロボットを用いて行うようにしてもかまわない。また、ユーザ107は電子メールを用いて検索条件を本システムへ登録するものとして述べるが、インターネットを介して本システムへ登録するようにしてもかまわない。さらに、本システムから前記検索条件に基づいて検索された前記テキストが電子メールを用いて該当ユーザに配送されるものとして述べるが、インターネット等を介して提示するようにしてもかまわない。

【0071】主メモリ104には、システム制御プログラム110、検索条件登録制御プログラム111、テキスト検索配送制御プログラム112、電子メールプログラム113、有限オートマトン114、ユーザリスト115、検索ターム重みテーブル116、適合度管理テーブル117、ワークエリア118、検索条件取得プログラム120、検索ターム抽出プログラム121、検索ターム重み算出プログラム122、検索用オートマトン作

成プログラム123、配送閾値設定プログラム124、テキスト取得プログラム130、テキスト検索プログラム131およびテキスト成形プログラム132が確保される。

【0072】検索用オートマトン作成プログラム123は、有限オートマトン作成プログラム140およびユーザリスト作成プログラム141で構成される。

【0073】なお、本実施例では、テキストから検索タームを抽出する手法として有限オートマトンを用いているが、本手法は有限オートマトンに限られるものではなく、有限オートマトンを用いてもよいし、NIKKEI BYTE, August 1987, p.p.175-189に記載されている拡張BM法（以下、従来技術3と呼ぶ）を用いてもよい。この場合、検索用オートマトン作成プログラム、有限オートマトン作成プログラムという表現は適当でなく、より一般化した表現を用いると、それぞれ検索用文字列照合テーブル作成プログラム、多重文字列照合テーブル作成プログラムの様になる。

【0074】テキスト検索プログラム131は、テキスト走査プログラム150、適合度算出プログラム151およびテキスト配送判定プログラム152で構成される。

【0075】以上のプログラムはハードディスク装置（本図には示していない）、フロッピーディスク（本図には示していない）等のコンピュータで読み書きできる記憶媒体に格納することもできる。

【0076】システム制御プログラム110は文書検索配送システム管理者によるキーボード101からの指示を受け起動する。

【0077】検索条件登録制御プログラム111およびテキスト検索配送制御プログラム112は、ユーザ107からの検索条件の登録指示やニュース配信元106からのテキストの配信により、システム制御プログラム110によって起動され、それぞれ検索条件取得プログラム120、検索ターム抽出プログラム121、検索ターム重み算出プログラム122および検索用オートマトン作成プログラム123の制御と、テキスト取得プログラム130、テキスト検索プログラム131およびテキスト成形プログラム132の制御を行う。

【0078】電子メールプログラム113にはワークステーション等で一般的に用いられている既存のメールプログラムを用いる。本電子メールプログラム113は、テキスト検索配送制御プログラム112の処理結果に応じて、システム制御プログラム110によって起動される。

【0079】以下、本実施例における文書検索配送システムの処理手順について説明する。

【0080】まず、システム制御プログラム110の処理手順を図3のPAD（Problem Analysis Diagram）図を用いて説明する。

【0081】システム制御プログラム110では、まずステップ300で、キーボード101から終了コマンドが入力されるまで、以下のステップを繰り返す。

【0082】この繰り返し処理では、まずステップ301でユーザ107から電子メールによって検索条件が送られてきているか否かを調べる。ここで、検索条件が送られてきている場合には、ステップ303で検索条件登録制御プログラム111を起動し、検索条件の作成登録を行う。

【0083】次に、ステップ302でニュース配信元106から電子メールによってテキストが送られてきているか否かを調べる。ここで、テキストが送られてきている場合には、ステップ304でテキスト検索配送プログラム112を起動し、テキストの検索を行う。

【0084】そして、ステップ305で、テキスト検索配送プログラム112におけるテキスト検索の結果を調べ、成立している検索条件が一つでも存在すると判断された場合には、ステップ306で電子メールプログラム113を起動し、成立した検索条件を指定したユーザに対して、該当テキストを電子メールを用いて配送する。

【0085】以上が、システム制御プログラム110の処理手順である。

【0086】次に図3に示したステップ303で起動される検索条件登録制御プログラム111の処理手順を図4のPAD図を用いて説明する。

【0087】検索条件登録制御プログラム111では、まずステップ400において、検索条件取得プログラム120を起動し、ユーザ107から電子メールによって送付された検索条件を取得し、ワークエリア118へ格納する。

【0088】次に、ステップ401において、検索ターム抽出プログラム121を起動し、ワークエリア118に格納されている検索条件中の種文書から検索タームを抽出し、ワークエリア118へ格納する。

【0089】次に、ステップ402において、検索ターム重み算出プログラム122を起動し、ワークエリア118に格納されている検索タームの重みを算出し、検索ターム重みテーブル116へ格納する。

【0090】次に、ステップ403において、検索用オートマトン作成プログラム123を起動し、検索条件中に含まれる全ての検索タームを照合する有限オートマトン114を作成する。

【0091】そして、ステップ404において、配送閾値設定プログラム124を起動し、検索条件中で指定された配送閾値を適合度管理テーブル117に格納する。

【0092】以上が、検索条件登録制御プログラム111の処理手順である。

【0093】次に、図4に示したステップ403で検索条件登録制御プログラム111により起動される検索用オートマトン作成プログラム123の処理手順を図6の

PAD図を用いて説明する。

【0094】検索用オートマトン作成プログラム123では、まずステップ600において、有限オートマトン作成プログラム140を起動し、検索ターム抽出プログラム121で抽出され、ワークエリア118に格納されている全ての検索タームを照合する有限オートマトン114を作成する。

【0095】次に、ステップ601において、ユーザリスト作成プログラム141を起動し、検索条件を指定したユーザ107の識別番号をリストとしてつなぎ、ユーザリスト115を作成する。

【0096】そして、ステップ602において、作成されたユーザリスト115をポインタを介して有限オートマトン114の該当する出力テーブルに接続する。

【0097】以上が、検索用オートマトン作成プログラム123の処理手順である。

【0098】以下に、図4に示した検索条件登録制御プログラム111の処理の流れを図10を用いて具体的に説明する。

【0099】まず、検索ターム抽出ステップ1000において、ユーザ107から電子メールによって送付された検索条件1010から検索ターム1011を抽出する。本図に示した例では、ユーザ1が登録した検索条件ユーザ1：“文書検索システムとしては、HiRetrievalがよく知られて・・・”からは、“文書”、“検索”、“システム”および“HiRetrieval”の4つの検索タームが抽出される。同様に、ユーザ2が登録した検索条件からは、“冬山”、“登山”、“防寒”および“対策”の4つの検索タームが抽出され、ユーザ3が登録した検索条件からは、“システム”、“文書”、“登録”、“性能”および“評価”の5つの検索タームが抽出される。

【0100】この検索ターム抽出技術としては、例えば、「特開平8-335222号公報」に開示されるように、形態素解析により単語辞書を参照し、種文書中に含まれる単語を検索タームとして抽出してもよいが、単語辞書に掲載されていない単語は抽出されないため、従来技術2に開示されるように、テキストデータベース中の統計情報を用いることにより、単語辞書を用いずに文書に記述された全ての単語を検索タームとして抽出するのが望ましい。なお、従来技術2ではテキストデータベースにおける各n-gramの出現確率を用いているが、文書検索配送システムの場合には、配信された全てのテキストにおける各n-gramの出現確率を用いるものとしても構わないし、配信されたテキストを登録したテキストデータベースにおける各n-gramの出現確率を用いてもよいし、あるいは配信されたテキスト以外の文書を登録したテキストデータベースにおける各n-gramの出現確率を用いるものとしてもよい。

【0101】次に、ステップ1001において、検索条

件1010から抽出された全ての検索ターム1011に対して、所定の算出式に基づいて重要度を算出し、検索ターム重みテーブル116に格納する。この重要度算出式としては、前述のIDF算出式(1)を用いてもよい。本算出式(1)を用いることにより、例えば、10万件のテキストが登録されているテキストデータベースにおいて、検索ターム“文書”が2,000件のテキストに出現している場合には、“文書”の重要度は“6.6”と算出される。ここで、テキストデータベースに登録されている全文書数および検索タームの出現している文書数は、それぞれニュース配信元106から配信されたテキスト数および該検索タームが出現したテキスト数としてもよいし、配信されたテキストとは異なるテキストが登録されているテキストデータベースを参照することにより求めてもよい。

【0102】次に、ステップ1002において、検索条件1010から抽出された全ての検索ターム1011を照合する有限オートマトン114を生成する。この有限オートマトンに対する検索タームの登録実施に関しては、従来技術1で開示されている方法を利用することができる。

【0103】以上が、図4に示した検索条件登録制御プログラム111の具体的な処理手順である。

【0104】次に図3に示したステップ304で、システム制御プログラム110により起動されるテキスト検索配送制御プログラム112の処理手順を図5のPAD図を用いて説明する。

【0105】テキスト検索配送制御プログラム112では、まずステップ500において、テキスト取得プログラム130を起動し、ニュース配信元106から電子メール等により配信されたテキストをワークエリア118へ格納する。

【0106】次に、ステップ501において、テキスト検索プログラム131を起動し、ワークエリア118に格納されたテキストを検索する。

【0107】次に、ステップ502において、所定の配送閾値を超える検索条件が一つ以上存在するか否かを判断し、存在する場合にはステップ503を実行する。

【0108】ステップ503では、テキスト成形プログラム132を起動し、ワークエリア118に格納されているテキストを電子メールプログラム113が配送できる形式に成形する。

【0109】以上が、テキスト検索配送制御プログラム112の処理手順である。

【0110】次に、図5に示したステップ501で起動されるテキスト検索プログラム131の処理手順を、図7のPAD図を用いて説明する。

【0111】テキスト検索プログラム131では、まずステップ700において、有限オートマトン114の出力テーブルに設けた検索ターム出現回数格納領域を

“0”にリセットする。

【0112】そして、ステップ701において、テキスト走査プログラム150を起動し、テキスト取得プログラム130によりワークエリア118に格納されたテキストを、有限オートマトン114で走査し、該テキスト中に一致する部分文字列が現れた検索タームの出現回数を計数する。

【0113】次に、ステップ702において、適合度算出プログラム151を起動し、上記テキスト走査プログラム150により計数された該テキスト内の検索タームの出現回数と、上記検索ターム重み算出プログラム122により検索ターム重みテーブル116に格納された各検索タームの重みを用いて、ユーザ107の登録した検索条件に対する該テキストの適合度を所定の適合度算出式を用いて算出し、適合度管理テーブル117へ格納する。

【0114】そして、ステップ703において、テキスト配送判定プログラム152を起動し、該テキストの適合度が適合度管理テーブル117に格納された配送閾値を超える検索条件を指定したユーザの識別子をテキスト検索配送制御プログラム112へ出力する。

【0115】以上が、テキスト検索プログラム131の処理手順である。

【0116】以下に、図7に示したテキスト検索プログラム131の処理の流れを図11を用いて具体的に説明する。

【0117】まず、テキスト取得プログラム130によりワークエリア118に格納されているテキスト1100に対して、ユーザ107により登録された全ての検索条件から抽出された検索タームを有限オートマトン114を用いて照合し、該テキスト内における出現回数を計数する（ステップ1100）。

【0118】本図に示した例は、テキスト：“膨大な文書の中から目的の文書を検索する。”を図10に示した有限オートマトン114を用いて走査し、適合する検索条件を検索する場合の例である。本図に示した有限オートマトン114では、末尾の状態に各検索タームのテキスト中での出現回数が付記されている。本例では、検索ターム“文書”に一致する部分文字列がテキスト中に2回出現しているため、その末尾の状態である状態2に“2”が付記され、検索ターム“検索”に一致する部分文字列がテキスト中に1回出現しているため、その末尾の状態である状態4に“1”が付記される。

【0119】そして、ステップ1101において、テキスト中の部分文字列と一致した全ての検索タームに対して、その末尾の状態に接続されたユーザリストを参照し、ユーザの登録した各検索条件に対するテキストの適合度を算出する。本実施例では、適合度の算出に前述の適合度算出式（2）を用いるが、他の方法を用いてもよい。この適合度算出式（2）を用いて、前述のテキス

ト：“膨大な文書の中から目的の文書を検索する。”の各ユーザの登録した検索条件に対する適合度を算出すると次のようになる。

【0120】

ユーザ1：9.2

ユーザ2：0

ユーザ3：6.6

そして、ステップ1102において、該適合度が適合度管理テーブル117に格納された配送閾値を超えているかを判別し、条件を満たしている場合には該当するユーザ識別子がテキスト検索配送制御プログラム112へ出力される。本実施例では、ユーザ1の配送閾値を“8.0”、ユーザ2の配送閾値を“3.5”、ユーザ3の配送閾値を“7.0”とするが、他のテキスト配送の条件を設定してもよい。

【0121】このテキスト配送の条件を用いて、各ユーザにより登録された検索条件に対する該テキストの適合度をみると、ユーザ1により登録された検索条件：“文書検索システムとしては、HiRetrievalがよく知られて・・・”の適合度がユーザ1の配送閾値“8.0”を超えているため、“ユーザ1”というユーザ識別子1111がテキスト検索配送制御プログラム112へ出力される。

【0122】以上説明したように、本実施例では、テキストのただ一度の走査で複数ユーザの検索条件に対する適合度を算出することが可能であり、またユーザ毎に配送閾値との比較を行うため、多くの情報を要求するユーザに対しては適合度の低いテキストまでも配送することが可能となり、重要な情報だけを要求するユーザに対しては適合度の高いテキストを配送することが可能となる。

【0123】なお、本実施例では、ユーザ毎に配送閾値を設定できるものとして説明したが、システムで共通の配送閾値をもつようにしてもよい。これにより、適合度管理テーブル117で必要となる記憶容量を削減することができる。

【0124】第一の実施例における適合度算出プログラム151は、有限オートマトン114の出力テーブルに接続されたユーザリスト115を順番にたどることにより、各ユーザの適合度を算出するため、ユーザ数の増加に伴い、全てのユーザに対する適合度算出終了までに要する時間が長大になるという問題がある。例えば、ユーザリストに接続された1個のユーザ識別子の処理が0.01秒で済むとしても、1万個のユーザ識別子の処理には100秒、すなわち最後のユーザの適合度算出が終了するまでには1分40秒も掛かってしまうことになる。

【0125】この問題を解決するために、本発明を適用した文書検索配送システムの第二の実施例では、ユーザ識別子に対して配送の優先度をつけ、優先度の高いユーザから適合度算出処理を行う。これにより、優先度の高いユーザに対して即時性の高い配送を行なうようにする

ものである。

【0126】本実施例は、第一の実施例（図1）とほぼ同様の構成を取るが、テキスト検索プログラム131aの処理手順と有限オートマトン114へのユーザリスト115の接続方法が異なる。本実施例におけるテキスト検索プログラム131aには、図13のPAD図に示すように、ステップ1300が追加される。また、有限オートマトン114には、図12に示すように、優先度識別子1200を介してユーザリスト115が接続される。

【0127】以下、第二の実施例におけるテキスト検索プログラム131aの処理手順を図13のPAD図を用いて説明する。

【0128】テキスト検索プログラム131aでは、まずステップ700において、有限オートマトン114の出力テーブルに設けた検索ターム出現回数格納領域を“0”にリセットする。

【0129】そして、ステップ701において、テキスト走査プログラム150を起動し、テキスト取得プログラム130によりワークエリア118に格納されたテキストを、有限オートマトン114で走査し、該テキスト中に一致する部分文字列が現れた検索タームの出現回数を計数する。

【0130】次に、ステップ1300において、有限オートマトン114に接続された優先度識別子1200の示す優先度の降順にステップ702とステップ703を繰り返し実行する。

【0131】ステップ702では、適合度算出プログラム151を起動し、前述のテキスト走査プログラム150により計数された該テキスト内の検索タームの出現回数と、前述の検索ターム重み算出プログラム122により検索ターム重みテーブル116に格納された各検索タームの重みを用いて、ユーザ107の登録した検索条件に対する該テキストの適合度を所定の適合度算出式を用いて算出し、適合度管理テーブル117へ格納する。

【0132】そして、ステップ703において、テキスト配送判定プログラム152を起動し、該テキストの適合度が適合度管理テーブル117に格納された配送閾値を超える検索条件を指定したユーザの識別子をテキスト検索配送制御プログラム112へ出力する。

【0133】以上が、テキスト検索プログラム131aの処理手順である。

【0134】以下、第二の実施例におけるテキスト検索プログラム131aの具体的な処理手順を図14に示す具体例で説明する。

【0135】テキスト検索プログラム131aでは、まず、ステップ1100において、前述のテキスト取得プログラム130によりワークエリア118に格納されているテキスト1110内に出現する検索タームの出現回数を計数する。本図に示した例では、テキスト111

0：“膨大な文書の中から目的の文書を検索する。”には、検索ターム“文書”が2回、検索ターム“検索”が1回出現しているという出現回数計数結果1410が得られる。

【0136】次に、ステップ1400において、図12に示した優先度識別子1200の“特急”に接続されたユーザリストを参照し、テキスト1110の適合度を算出する。図14の例では、ユーザ1の登録した検索条件に対する適合度が“9.2”と算出される。

【0137】そして、ステップ1401において、適合度管理テーブルに格納された配送閾値を超えているかを判別し、条件を満たしている場合にはユーザ識別子がテキスト検索配送制御プログラム112へ出力される。本実施例では、ユーザ1の配送閾値を“8.0”とするが、他のテキスト配送の条件を設定してもよい。ユーザ1の登録した検索条件に対する適合度“9.2”は配送閾値“8.0”を超えているので、ユーザ識別子“ユーザ1”が出力される。

【0138】次に、ステップ1402において、図12に示した優先度識別子1200の“普通”に接続されたユーザリストを参照し、テキスト1110の適合度を算出する。図14の例では、ユーザ2およびユーザ3の登録した検索条件に対する適合度が、それぞれ“0”、“6.6”と算出される。

【0139】そして、ステップ1403において、適合度管理テーブルに格納された配送閾値を超えているかを判別し、条件を満たしている場合にはユーザ識別子がテキスト検索配送制御プログラム112へ出力される。この結果、ユーザ2およびユーザ3が登録した検索条件に対する適合度は、それぞれの配送閾値を超えていないので、ユーザ識別子は出力されない。

【0140】以上説明したように、本実施例によれば優先度の高いユーザに対して、テキストの検索を優先的に行なうことができるようになる。このため、ユーザ数が増加した場合においても、優先度の高いユーザに対して即時性の高い文書検索配送システムを提供することが可能となる。

【0141】なお、本実施例では、優先度識別子1200には“特急”と“普通”というユーザが設定した優先度に基づいた配信優先度の種別を持つものとして説明したが、例えば、“部長”や“課長”等の役職に基づいた配信優先度を持つものとしてもよいし、“正規ユーザ”や“体験ユーザ”等の契約料金に基づいた配信優先度を持つものとしてもよい。

【0142】次に、本発明の第三の実施例について図15を用いて説明する。

【0143】第一、第二の実施例においては、種文書から抽出された検索タームを種文書の種類によらず同一の重要度を持つものとして適合度を算出するものとしたが、種文書に記述されている主題が異なる場合でも同一

の重みを持つため、種文書の主題を的確に反映することができないという問題がある。

【0144】例えば、検索条件“文書検索システムとしては、HiRetrievalがよく知られて・・・”から抽出された検索ターム“HiRetrieval”は、「文書検索システム」の一例として挙げられているのに対し、検索条件“HiRetrievalでは、ANDやOR等の論理演算を行なうことができる。また、HiRetrievalではSGMLやXML等で記述された構造化文書を登録できる。さらには、HiRetrievalでは・・・”から抽出された検索ターム“HiRetrieval”は、文書の主題を表わす単語であり、重要度が高い。

【0145】本発明を適用した文書検索配送システムの第三の実施例では、ユーザリストに各検索条件における検索ターム重みを付加することにより、上記問題を解決するものである。

【0146】本実施例は、第一の実施例（図1）とほぼ同様の構成を取るが、検索条件登録制御プログラム111が異なり、図15に示すように検索条件重み算出プログラム1500が追加される。また、ユーザリスト作成プログラム141aで出力されるユーザリスト115の形式が異なるとともに、適合度算出プログラム151の処理手順が異なる。

【0147】以下、第一の実施例とは異なる検索条件登録制御プログラム111aの処理手順について図16を用いて説明する。

【0148】検索条件登録制御プログラム111aでは、まずステップ400において、検索条件取得プログラム120を起動し、ユーザ107から電子メールによって送付された検索条件を取得し、ワークエリア118へ格納する。

【0149】次に、ステップ401において、検索ターム抽出プログラム121を起動し、ワークエリア118に格納されている検索条件中の種文書から検索タームを抽出し、ワークエリア118へ格納する。

【0150】次に、ステップ402において、検索ターム重み算出プログラム122を起動し、ワークエリア118に格納されている検索タームの重みを算出し、検索ターム重みテーブル116へ格納する。

【0151】次に、ステップ1600において、検索条件重み算出プログラム1500を起動し、ワークエリア118に格納されている各検索タームの検索条件毎の重みを算出し、ワークエリア118へ格納する。

【0152】次に、ステップ1601において、検索用オートマトン作成プログラム123aを起動し、検索条件中に含まれる全ての検索タームを照合する有限オートマトン114を作成する。

【0153】そして、ステップ404において、配送閾値設定プログラム124を起動し、検索条件中で指定された配送閾値を適合度管理テーブル117に格納する。

【0154】以上が、検索条件登録制御プログラム111aの処理手順である。

【0155】次に、検索条件登録制御プログラム111aのステップ1601で起動される検索用オートマトン作成プログラム123aの処理手順を図17に示すPAD図を用いて説明する。

【0156】検索用オートマトン作成プログラム123aでは、まずステップ600において、有限オートマトン作成プログラム140を起動し、検索ターム抽出プログラム121で抽出され、ワークエリア118に格納されている全ての検索タームを照合する有限オートマトン114を作成する。

【0157】次に、ステップ1700において、ユーザリスト作成プログラム141aを起動し、検索条件を指定したユーザ107の識別番号と上記検索条件重み算出プログラム1500によりワークエリア118に格納されている各検索タームの検索条件毎の重みをリストとしてつなぎ、ユーザリスト115aを作成する。

【0158】そして、ステップ1701において、作成されたユーザリスト115aをポインタを介して有限オートマトン114の該当する出力テーブルに接続する。

【0159】以上が、検索用オートマトン作成プログラム123aの処理手順である。

【0160】以下に、図15に示した第三の実施例における検索条件登録制御プログラム111aの処理の流れを図18を用いて具体的に説明する。

【0161】まず、検索ターム抽出ステップ1000において、ユーザ107から電子メールによって送付された検索条件1010から検索ターム1011を抽出する。本図に示した例では、ユーザ1が登録した検索条件ユーザ1：“文書検索システムとしては、HiRetrievalがよく知られて・・・”からは、“文書”、“検索”、“システム”および“HiRetrieval”の4つの検索タームが抽出される。同様に、ユーザ2が登録した検索条件からは、“冬山”、“登山”、“防寒”および“対策”の4つの検索タームが抽出され、ユーザ3が登録した検索条件からは、“システム”、“文書”、“登録”、“性能”および“評価”の5つの検索タームが抽出される。

【0162】この検索ターム抽出技術としては、例えば、「特開平8-335222号公報」に開示されるように、形態素解析により単語辞書を参照し、種文書に含まれる単語を検索タームとして抽出してもよいが、単語辞書に掲載されていない単語は抽出されないため、従来技術2に開示されるように、テキストデータベース中の統計情報を用いることにより、単語辞書を用いずに文書に記述された全ての単語を検索タームとして抽出するのが望ましい。なお、従来技術2ではテキストデータベースにおける各n-gramの出現確率を用いているが、文書検索配送システムの場合には、配信された全てのテキス

トにおける各n-gramの出現確率を用いるものとしても構わないし、配信されたテキストを登録したテキストデータベースにおける各n-gramの出現確率を用いてもよいし、あるいは配信されたテキスト以外の文書を登録したテキストデータベースにおける各n-gramの出現確率を用いるものとしてもよい。

【0163】次に、ステップ1001において、検索条件1010から抽出された全ての検索ターム1011に対して、所定の算出式に基づいて検索ターム重要度を算出し、検索ターム重みテーブル116に格納する。この検索ターム重要度の算出式としては、前述のIDF算出式

(1)を用いてもよい。例えば、10万件のテキストが登録されているテキストデータベースにおいて、検索ターム“文書”が2,000件のテキストに出現している場合には、“文書”の重要度は“6.6”と算出される。ここで、テキストデータベースに登録されている全文書数および検索タームの出現している文書数は、それぞれニュース配信元106から配信されたテキスト数および該検索タームが出現したテキスト数としてもよいし、配信されたテキストとは異なるテキストが登録されているテキストデータベースを参照することにより求めてもよい。

【0164】次に、ステップ1800において、検索条件1010から抽出された全ての検索ターム1011に対して、所定の算出式に基づいて各検索条件における検索条件重要度を算出し、ワークエリア118に格納する。この検索条件重要度としては、例えば、検索条件における出現回数そのものとしてもよい。

【0165】次に、ステップ1002において、検索条件1010から抽出された全ての検索ターム1011を照合する有限オートマトン114を生成する。この有限オートマトンに対する検索タームの登録実施に関しては、従来技術1で開示されている方法を利用することができる。

【0166】以上が、図15に示した検索条件登録制御プログラム111aの具体的な処理手順である。

【0167】次に、本発明の第三の実施例において、第一、第二の実施例とは異なる適合度算出プログラム151aの処理手順を図19に示すPAD図を用いて説明する。

【0168】適合度算出プログラム151aは、テキスト走査プログラム150により照合された全ての検索タームについて、ステップ1901～1905を繰り返し実行する(ステップ1900)。

【0169】まず、ステップ1901において、テキスト走査プログラム150により計数された検索タームの出現回数を取得する。そして、ステップ1902において、検索ターム重みテーブル116に格納されている検索タームの重みを取得する。

【0170】次に、ユーザリスト115aの末尾までの各ユーザ識別子に対して、ステップ1904～1905

を繰り返し実行する。まず、ステップ1904において、ユーザ識別子と各検索条件における検索条件重みを取得する。そして、ステップ1905において、所定の算出式に基づいて各検索条件に対する適合度を算出する。

【0171】以上が、適合度算出プログラム151aの処理手順である。

【0172】以上説明したように、本実施例によれば、各ユーザが登録した検索条件の主題を表わす検索タームに高い重みを付加することができ、精度の高い文書検索配送システムを提供することが可能となる。

【0173】次に、本発明の第四の実施例について説明する。

【0174】以上説明した、第一、第二、第三の実施例では、図11のステップ1102において、適合度があらかじめ設定された所定の値(以下配送閾値と呼ぶ)以上であるという条件を満たしているテキストを配送するものとしている。しかし、この方式では、各ユーザの設定した検索条件に対して、適切な配送閾値を設定することが困難であるという問題点がある。例えば、設定された配送閾値が高すぎると、ユーザには所望のテキストが配送されなくなってしまう。また、設定された配送閾値が低すぎると、ユーザにとって必要でないテキストも大量に配送されてしまうという問題がある。即ち、初期設定された配送閾値を変更する必要がある場合、ユーザにとって、どの程度配送閾値を変更すればノイズなく、また漏れのない配送結果を得られるかがわからない、といった問題がある。

【0175】この問題を解決するために、本発明を適用した文書検索配送システムの第四の実施例では、ユーザに対して、適切な配送閾値を設定するために有用な情報(以下、配送閾値設定支援情報と呼ぶ)を提示する。さらに、ユーザが仮に設定した配送閾値に対して、算出された適合度がその配送閾値を上回る過去のテキストを表示(以下、配送閾値の設定の試行と呼ぶ)する。以上の情報により、ユーザが各自好適な配送閾値を設定することを可能にする。

【0176】次に、本発明の第四の実施例のシステム構成を図21に示す。

【0177】本実施例は、図1に示す第一の実施例とほぼ同様の構成をとり、図21に示すように配送閾値更新プログラム2000、配送閾値設定支援情報提示プログラム2001、および配送閾値設定試行プログラム2002が追加される。また、個人適合度判定情報管理エリア2003とテキスト保存エリア2004が追加される。

【0178】個人適合度判定情報管理エリア2003には、過去に受信したテキストに対して算出された、各ユーザの検索条件に対する適合度や、各ユーザに配送したかを示すフラグが格納されている。ま

た、テキスト保存エリア 2004 には、過去に受信したテキストの内容と受信時刻が格納されている。

【0179】本実施例では、個人適合度判定情報管理エリア 2003 およびテキスト保存エリア 2004 に格納された情報を用いて、配送閾値支援情報提示プログラム 2001 が配送閾値設定支援情報をユーザに提示する。また、同様に個人適合度判定情報管理エリア 2003 およびテキスト保存エリア 2004 に格納された情報を用いて、配送閾値設定試行プログラム 2002 がユーザに配送閾値の設定の試行機能を提供する。ユーザはこれらの機能により適切な配送閾値を決定することが可能になり、配送閾値更新プログラム 2000 によって、決定した配送閾値をシステムに登録する。ここで登録された配送閾値は、テキスト検索配送制御プログラム 112a により、受信したテキストを各ユーザに配送するか否かを判定する際に用いられる。

【0180】以下、本実施例ではユーザからの配送閾値設定支援情報の提示要求、配送閾値設定の要求、配送閾値設定試行の要求は電子メールの形で送付されるものとして説明するが、Webブラウザなどの他のネットワークアプリケーションを介して本システムに要求されるものとしてもかまわない。さらに、前記のユーザの要求に基づいた配送閾値設定支援情報、配送閾値設定試行結果は、本システムから電子メールを用いて該当ユーザに配送されるものとして述べるが、Webブラウザなどの他のネットワークアプリケーションを介して提示されるようにしてもかまわない。

【0181】以下、第四の実施例における、各プログラムの処理手順を説明する。

【0182】まず、第四の実施例におけるシステム制御プログラム 110a の処理手順を図 22 の PAD 図を用いて説明する。

【0183】本実施例におけるシステム制御プログラム 110a の処理手順は、第一の実施例におけるシステム制御プログラム 110 に、ステップ 2100～ステップ 2105 の処理が加わったものである。

【0184】ステップ 300 の繰り返し処理において、ステップ 301、ステップ 302 の処理の後、ステップ 2100 で、ユーザから配送閾値が送られているかを調べる。ここで配送閾値が送られてきている場合には、ステップ 2103 で配送閾値更新プログラム 2000 を起動し、該ユーザの配送閾値を設定する。

【0185】次に、ステップ 2101 で、ユーザ 107 から配送閾値設定支援情報の提示要求が送られてきているかを調べる。ここで、配送閾値設定支援情報の提示要求が送られてきている場合には、ステップ 2104 で配送閾値設定支援情報提示プログラム 2001 を起動し、配送閾値設定支援情報を該ユーザに送信する。

【0186】次にステップ 2102 で、ユーザから配送閾値設定の試行要求が送られてきているかを調べる。こ

こで、配送閾値の試行要求が送られてきている場合には、ステップ 2105 で配送閾値設定試行プログラム 2002 を起動し、配送閾値設定の試行を行う。

【0187】以上が、システム制御プログラム 110a の処理手順である。

【0188】以下、図 22 に示したステップ 2103 においてシステム制御プログラムにより起動される、配送閾値更新プログラム 2000 の処理手順を図 23 の PAD 図を用いて説明する。

【0189】配送閾値更新プログラム 2000 では、まずステップ 2200 において、ユーザ 107 から電子メールによって送付された配送閾値を取得する。

【0190】次に、ステップ 2201 において、ユーザリスト 115b に格納された該ユーザの配送閾値を、ユーザ 107 から指定された配送閾値に更新する。ここで、ユーザリスト 115b 中の各ユーザの配送閾値の初期設定値は、管理者が予め定めた値であってもよいし、ユーザ 107 が検索条件に登録するときに入力するものとしてもよい。

【0191】以上が、配送閾値更新プログラム 2000 の処理手順である。

【0192】以下、図 22 に示したステップ 2104 においてシステム制御プログラムにより起動される、配送閾値設定支援情報提示プログラム 2001 の処理手順を、図 24 の PAD 図を用いて説明する。

【0193】本プログラムは、過去に各ユーザに対し行った、配送判定の履歴をもとに、適切な閾値を設定するための判断材料となる情報を提供するものである。

【0194】配送閾値設定支援情報提示プログラム 2001 では、まずステップ 2300 において、配送閾値設定支援情報の提示を要求したユーザ 107 のユーザ識別子をもとに、個人適合度判定情報管理エリア 2003 から、過去所定の期間内にニュース配信元 106 から受信したテキストに関する、該ユーザの個人適合度判定情報を読み込む。ここで個人適合度判定情報とは、ニュース配信元 106 から受信した全てのテキストについての、各ユーザに対して算出した適合度、各ユーザに配送したかどうかを示したフラグ、といったデータのことであり、具体例は後に説明する。

【0195】次に、ステップ 2301 において、テキスト保存エリア 2004 から、過去所定の期間内にニュース配信元 106 から受信したテキストの内容を読み込む。

【0196】次に、ステップ 2302 において、ステップ 2301 で読み込んだ情報のうち、該ユーザに配送したテキストに関するものを一覧にして、ワークエリア 118 に出力する。

【0197】次に、ステップ 2303 において、該ユーザの検索条件に対して算出した適合度別のテキスト数、および、適合度を縦軸にとりテキストの受信時刻を横軸

にとったグラフ（以下、適合度の分布情報と呼ぶ）を作成し、ワークエリア 118 に出力する。ここで、グラフの縦軸や横軸にとる変数は、ステップ 2301、2302 において読み込んだその他の情報でも構わない。

【0198】次に、ステップ 2304 において、ワークエリア 118 に格納されている情報を、電子メールプログラム 113 が配送できる形式に成形する。

【0199】ワークエリア 118 に格納された情報は、電子メールプログラム 113 により、該ユーザに配送される。

【0200】以上が、配送閾値設定支援情報提示プログラム 2001 の処理手順である。

【0201】以下、図 24 に示した配送閾値設定支援情報提示プログラム 2001 の処理の流れを、図 25 を用いて具体的に説明する。

【0202】まず、ステップ 2300 において、個人適合度判定情報管理エリア 2003 から、配送閾値設定支援情報の提示を要求してきたユーザ 107 のユーザ識別子に対応する、過去所定の期間内（ここでは仮に過去 24 時間とする）の個人適合度判定情報 2400 を取得する。本図に示した例では、ユーザ 1 に関するテキスト 1、テキスト 2、テキスト 3 のテキスト識別子、算出された適合度、配信したかしなかったかを示すフラグ、配信判定を行った時点でユーザ 107 が設定していた配送閾値、を取得する。

【0203】次に、ステップ 2301 において、テキスト保存エリア 2004 から、過去 24 時間内にニュース配信元 106 から配送されたテキストの内容を取得する。本図に示した例では、テキスト 1、テキスト 2、テキスト 3 の内容が取得される。

【0204】本実施例では、ステップ 2300 およびステップ 2301 において、過去 24 時間にニュース配信元 106 から配信されたテキストを対象とするが、他の期間にしてもよいし、保存されている全てを対象としてもよいし、またはユーザ 107 が期間を指定できるようにしてもよい。

【0205】次に、ステップ 2302 において、ステップ 2300 とステップ 2301 で取得した情報を、テキスト識別子をもとに照合し、該ユーザに配送したテキストに関する情報を抽出して、ワークエリア 118 に出力する。本図に示した例では、ユーザ 1 に配送されたテキスト 1 およびテキスト 3 の、テキスト識別子、適合度、当時設定していた配送閾値、配送時刻の一覧 2401 を、ワークエリア 118 に出力する。なお、ここで出力する情報は、ステップ 2300 およびステップ 2301 で取得した情報のうち任意の組み合わせを選らんで構わない。本実施例では適合度の順に出力するものとするが、テキストの配送時刻順に出力してもよいし、ユーザが適合度順か配送時刻順かを選択できるようにしてもよい。また、テキスト識別子のかわりに、文書の内容の最

初の一文を抽出したものなどを出力してもよいし、タイトルのような属性があれば、それを出力してもよい。

【0206】次に、ステップ 2303 において、ステップ 2300 とステップ 2301 で取得した情報を、テキスト識別子をもとに照合し、テキストの適合度の分布情報を作成し、ワークエリア 118 に追加出力する。本図に示した例では、過去 24 時間の適合度別のテキストの数 2402、および時刻を横軸にとり、適合度を縦軸にとったテキストの分布のグラフ 2403 を出力する。

【0207】このようなグラフを参照することによりユーザ 107 は、閾値をある値にすると、どの位の期間にどの位の量のテキストが配送されるようになるかが、視覚的に容易に判断できる。また、配信状況の時間推移を把握できる。例えば、図 26 に示すようなグラフが得られたときは「所望のテキスト（高い適合度のテキスト）がだんだんニュース配信元から配信されなくなってきた」といったことが判断できるので、そのときは配送閾値を低く修正することができる。さらに、所望のテキストが頻繁に配送されるような時間帯を知ることでもできる。例えば、図 27 に示すようなグラフが得られたときは、「おおよそ 18:00～21:00 の間に所望のテキストが多く配信されている」といったことが判断できる。これにより、たまたま所望のテキストがあまり配送されない時間帯に配送閾値を設定すると局所最適化されてしまい、配送閾値を適切な値よりも低く設定してしまう、といったことを避けることができる。

【0208】本ステップにおいて、本実施例では、過去 24 時間にニュース配信元 106 から配信された全てのテキストを対象としているが、該ユーザ 107 に配送しなかったテキストのみを対象としてもよい。また、該ユーザ 107 に配送したか否かの情報をグラフに追加してもよいし、該ユーザ 107 の設定していた配送閾値の時刻変化を同時に出力してもよい。また、本実施例では、適合度として、個人適合度判定情報管理エリア 2003 に格納されているものを用いているが、テキスト保存エリア 2004 に格納されているテキストを再び走査して、該ユーザの現時点での検索条件との適合度を再計算し、その結果を用いてもよい。

【0209】次に、ステップ 2304 において、ワークエリア 118 に格納されている情報を電子メールプログラム 113 が配送できる形に成形する。

【0210】以上が配送閾値設定支援情報提示プログラム 2001 の具体的な処理の流れである。本実施例では、配送閾値設定支援情報は、ユーザの要求により提示されるものとしたが、ある定まった時刻に全てのユーザ 107 に提示されるようにしてもよい。

【0211】以下、図 22 に示したステップ 2105 においてシステム制御プログラムにより起動される、配送閾値設定試行プログラム 2002 の処理手順を図 28 の PAD 図を用いて説明する。

【0212】本プログラムは、過去に受信したテキストに対しユーザ毎に算出した適合度をもとに、ユーザが示した配送閾値を超える適合度を算出された過去のテキストを提示するものである。

【0213】配送閾値設定試行プログラム2002では、まずステップ2500において、ユーザ107から電子メールによって送付された配送閾値を取得する。

【0214】次にステップ2501において、該ユーザのユーザ識別子をもとに、個人適合度判定情報管理エリア2003から、該ユーザの個人適合度判定情報のうち、ステップ2500において取得した配送閾値を超える適合度のテキストの情報を読み込む。

【0215】次に、ステップ2502において、ステップ2501において取得した適合度判定情報のテキスト識別子に対応するテキストの内容と受信時刻を、テキスト保存エリア2004から読み込み、適合度判定情報とともにワークエリア118に出力する。

【0216】次に、ステップ2503において、ワークエリア118に格納されている情報を、電子メールプログラム113が配送できる形式に成形する。

【0217】ワークエリア118に格納された情報は、電子メールプログラム113により、該ユーザに配送される。

【0218】以上が、配送閾値設定試行プログラム2002の処理手順である。

【0219】以下、図28に示した配送閾値設定試行プログラム2002の処理の流れを、図29を用いて具体的に説明する。

【0220】まず、ステップ2500において、ユーザ107から送付された配送閾値2600を取得する。

【0221】次に、ステップ2501において、個人適合度判定情報管理エリア2003から、ステップ2500において取得した配送閾値2600より高い適合度を算出されているテキストの情報2601を取得する。本図に示した例では、ユーザ識別子「ユーザ1」のユーザから配送閾値として9.0が送付されており、個人適合度情報管理エリア2003から9.0を上回る10.0の適合度を算出されたテキスト1の内容と受信時刻が取得される。なお、以下「ユーザ識別子「ユーザ1」のユーザ」という表記は、「ユーザ識別子」が「ユーザ1」であるユーザ、を示すこととする。本実施例では、本ステップで参照する適合度として、個人適合度判定情報管理エリア2003に格納されているものを用いているが、テキスト保存エリア2004に格納されているテキストを再び走査して、現在の該ユーザの検索条件との適合度を再計算し、その結果を用いてもよい。

【0222】次に、ステップ2502において、テキスト保存エリアに格納されているテキストの情報のうち、ステップ2501において取得したテキスト識別子に対応するテキスト情報2602、および適合度判定情報2

601を、ワークエリア118に出力する。本図に示した例では、ステップ2501において取得したテキスト1の、適合度、配送時刻、テキストの内容がワークエリア118に出力される。ここで、テキストの内容のかわりに、文書の内容の最初の一文を抽出したものなどを出力しても良いし、タイトルのような属性があれば、それを出力してもよい。

【0223】次に、ステップ2503において、ワークエリア118に格納されている情報を電子メールプログラム113が配送できる形に成形する。

【0224】ここで例えば、適合度判定情報管理エリア2003には、過去24時間にニュース配信元106から配信されたテキストに関する情報が保存されているものとし、ステップ2502においてテキストのタイトルが出力されるものとする。また、ユーザ107が24時間に2件のテキストの配送を希望しているとする。ここでユーザ107が、試行する配送閾値2600として「10.0」を送付すると、本図に示した例では、1件（テキスト1）のタイトルしか出力されない。そこで次に、試行する配送閾値2600として「6.0」を送付してみると、本テキスト1とテキスト3の2件のテキストのタイトルが出力される。その結果ユーザ107は、配送閾値更新プログラム2000により配送閾値を「6.0」に変更すれば、次の24時間で2件のテキストが配送されてきそうだと判断できる。

【0225】また、ステップ2502において、テキストのタイトルと共に内容の最初の一文を出力するものとする。このとき、試行する配送閾値2600を十分低くしてみると、適合度が設定していた配送閾値未満であったためユーザ107に配送されなかったテキストの、タイトルと最初の一文が出力される。ユーザ107はそれらを参照し、所望のテキストを見つけた場合には配送閾値更新プログラム2000により配送閾値をそのテキストの適合度よりも低く修正する。これにより、配送漏れのない配送閾値を設定することができる。

【0226】さらに、ステップ2502で出力される情報を、配送閾値設定支援情報提示プログラム2001において示したように、適合度を縦軸に、受信時刻を横軸にとったテキストの適合度の分布グラフとする。例えば図30に示すように、過去にニュース配信元106から配信されたテキストのうち、試行する配送閾値2600を上回るものを色分けしたり、記号を変えたりしてテキストの適合度の分布を表示する。

【0227】本図に示した例では、配送閾値を試行する値に設定していたならば、過去の24時間には5件のテキストが配送されてきていたことがわかる。このことから、この試行した値に配送閾値を設定すれば、次の24時間に同程度のテキストが配送されるであろうと判断できる。このようにユーザ107は、試行する配送閾値2600に配送閾値を変更するとどのような配送結果が得ら

れるかを視覚的に容易に知ることができ、この情報に基づき配送閾値の設定を行なうことができる。

【0228】以上が配送閾値設定試行プログラム2002の具体的な処理の流れである。

【0229】以下、図22に示したステップ304においてシステム制御プログラムにより起動される、テキスト検索配送制御プログラム112aの処理手順を図31のPAD図を用いて説明する。

【0230】本プログラムは、ニュース配信元106から送られるテキストに対し、ユーザ毎に検索条件との適合度を判定し、ユーザ毎に配送判定を行い、さらにテキストの内容およびユーザ毎の配送判定履歴を保存するものである。

【0231】テキスト検索配送制御プログラム112aでは、まずステップ2700において、テキスト取得プログラム130aを起動し、ニュース配信元106から電子メール等により配信されたテキストをワークエリア118へ格納する。さらに、テキストの内容と、受信時刻をテキスト保存エリア2004に格納する。

【0232】次に、ステップ2701において、テキスト検索プログラム131を起動し、ワークエリア118に格納されたテキストを検索し、各ユーザの設定した検索条件との適合度を算出する。さらに、各ユーザ毎に配送判定を行い、その結果を個人適合度判定情報管理エリア2003に格納する。

【0233】次に、ステップ2702において、所定の条件を満足する検索条件が一つ以上存在するか否かを判断し、存在する場合にはステップ2703を実行する。

【0234】ステップ2703では、テキスト成形プログラム132を起動し、ワークエリア118に格納されているテキストを電子メールプログラム113が配送できる形式に成形する。

【0235】以上が、テキスト検索配送制御プログラム112aの処理手順である。

【0236】以下、第四の実施例におけるテキスト検索配送制御プログラム112aの処理手順を、図32を用いて具体的に説明する。

【0237】まずステップ2700において、テキスト取得プログラム130aを起動し、ニュース配信元106から電子メール等により配信されたテキスト2810をワークエリア118に格納する。さらに、該テキストにテキスト識別子を付与し、テキスト保存エリア2004に、該テキストの内容および受信時刻を格納する。

【0238】次に、ステップ2800において、第一の実施例の図7で説明したものと同一手順でステップ700、ステップ701、およびステップ702を実行し、適合度算出を行った結果を、適合度管理テーブル117に格納する。

【0239】次に、ステップ2801において、適合度管理テーブル117に格納された適合度が、ユーザリス

ト115bに格納されている配送閾値よりも上回るユーザのユーザ識別子をテキスト検索配送制御プログラム112aへ出力する。さらに、テキスト識別子、算出した適合度、配送可否、現在の配送閾値、を各ユーザ毎の個人適合度判定情報管理エリア2003に格納する。本図に示した例では、ユーザ1に対して10.0という適合度が算出されたが、ユーザリスト115b中のユーザ1の配送閾値[12.0]を上回らないため、ユーザ1には該テキストは配送されない。ユーザ3に対しては6.6という適合度が算出されたが、ユーザリスト115b中のユーザ2の配送閾値[5.0]を上回るため、ユーザ3には該テキストが配送される。さらに、個人適合度判定情報管理エリア2003のユーザ1の部分にテキスト識別子[テキスト1]、適合度[10.0]、配送可否[NO]、現在の閾値[12.0]、といった情報が格納される。図32に示すように、ユーザ2、ユーザ3についても同様である。なお、これらの適合度判定情報は、先に説明したように配送閾値設定支援情報提示プログラム2001および配送閾値設定試行プログラム2002において用いられる。

【0240】次に、ステップ2703において、ワークエリア118に格納されている情報を電子メールプログラム113が配送できる形に成形する。

【0241】以上が、テキスト検索配送制御プログラム112aの処理手順である。

【0242】以上が、第四の実施例における、各プログラムの処理手順である。

【0243】以上示したように本実施例によれば、ニュース配信元から受信した全テキストや、各ユーザに対して行なった適合度算出の履歴を保存しておき、それらを配送閾値の設定時にユーザに提示する形態をとるため、各ユーザがそれらの情報を参考に適切な配送閾値を設定することが可能になる。したがって、初期設定された配送閾値を変更する必要がある場合、ユーザにとって、どの程度配送閾値を変更すればノイズなく、また漏れの無い配送結果を得られるかがわからない、といった問題が解決できる。

【0244】ここで、本実施例において説明した配送テキストの選択方法は、上述した有限オートマトンや拡張BM法といったテキストの検索方法に依るものではなく、ここで説明したもの以外の検索方法を用いたシステムでも同様に用いることができるものである。

【0245】また、本実施例において説明した配送テキストの選択方法は、検索条件に対するテキストの適合度算出方法によるものではなく、第一、第二、第三の実施例に示した適合度算出方法、またそれ以外の適合度算出方法でも同様に用いることができるものである。

【0246】次に、本発明の第五の実施例について説明する。

【0247】第四の実施例においては、適切な配送閾値

を設定するために、過去の適合度算出の履歴を参照する形態をとっているが、時々刻々変化するテキストの配信状況に完全には対応できない。例えば、ある時期に多量のテキストが配送されたため配送閾値を高く設定しても、その後は適合度がその配送閾値を超えるようなテキストの数が減少し、結果としてユーザにテキストが全く配送されなくなる可能性がある。このような場合ユーザは所望のテキストが出現しなかったのか、配送閾値を高く設定しすぎたのか判断できない、といった問題がある。

【0248】この問題を解決するために、本発明を適用した文書検索配送システムの第五の実施例では、所定の時間毎に、ユーザが設定した配送閾値を超える適合度のテキストに加えて、ユーザが設定した数のテキストを配送する（以下、追加配送とよぶ）ものである。

【0249】次に、本発明の第五の実施例のシステム構成を図33に示す。

【0250】本実施例は、図21に示す第四の実施例とほぼ同様の構成をとり、追加配送テキスト数設定プログラム2900、およびテキスト追加配送プログラム2901が追加される。

【0251】本実施例では、個人適合度判定情報管理エリア2003およびテキスト保存エリア2004に格納された情報を用いて、テキスト追加配送プログラム2901が、各ユーザにテキストを追加配送する。ここで追加配送するテキストの数は、ユーザが追加配送テキスト数設定プログラム2900を用いて設定する。

【0252】以下、本実施例では、ユーザからの追加配送テキスト数の設定要求は電子メールの形で送付されるものとして説明するが、Webブラウザなどの他のネットワークアプリケーションを介して本システムに要求されるものとしてもかまわない。さらに、前記のユーザの要求に基づいた、テキストの追加配送は本システムから電子メールを用いて行われるものとして述べるが、Webブラウザなどの他のネットワークアプリケーションを介して提示されるようにしてもかまわない。

【0253】以下に、第五の実施例における、各プログラムの処理手順について説明する。

【0254】まず、第五の実施例におけるシステム制御プログラム110bの処理手順を図34のPAD図を用いて説明する。

【0255】本実施例におけるシステム制御プログラム110bの処理手順は、第四の実施例におけるシステム制御プログラム110aに、ステップ3000～ステップ3003が追加されたものである。

【0256】ステップ3000では、ユーザ107から追加配送テキスト数の設定要求が送られてきているかを調べる。ここで、追加配送テキスト数の設定要求が送られてきている場合には、ステップ3002で追加配送テキスト数設定プログラム2900を起動し、該ユーザの

追加配送テキスト数を設定する。

【0257】次にステップ3001で、現在の時刻が所定の条件を満たす時刻であるかを判定する。もし、所定の条件を満たす時刻である場合には、ステップ3003においてテキスト追加配送プログラム2901を起動する。例えば、「毎日午前零時にテキスト追加配送プログラム2901を起動する」という条件を設定して、テキスト追加配送プログラム2901を起動することができる。

【0258】以上が、システム制御プログラム110bの処理手順である。

【0259】以下、図34に示したステップ3002においてシステム制御プログラム110bにより起動される、追加配送テキスト数設定プログラム2900の処理手順を図35のPAD図を用いて説明する。

【0260】追加配送テキスト数設定プログラム2900では、まずステップ3100において、ユーザ107から電子メールによって送付された、追加配送テキスト数を取得する。

【0261】次に、ステップ3101において、該ユーザのユーザ識別子をもとに、ユーザリスト115cにおける該ユーザの追加配送テキスト数を更新する。ここで、ユーザリスト115c中の各ユーザの追加配送テキスト数の初期設定値は、管理者が予め定めた値であってもよいし、ユーザ107が検索条件を登録するときに入力するものとしてもよい。

【0262】以上が、追加配送テキスト数設定プログラム2900の処理手順である。

【0263】以下、図34に示したステップ3003においてシステム制御プログラム110bにより起動される、テキスト追加配送プログラム2901の処理手順を、図36のPAD図を用いて説明する。

【0264】本プログラムは、各ユーザが要求する配送テキスト数を満たすように、適合度が配送閾値を超えなかったテキストを追加配送するものである。

【0265】テキスト追加配送プログラム2901は、ステップ3200において、ユーザリスト115cに格納された全てのユーザについて、ステップ3201～ステップ3204を繰り返し実行する。

【0266】まずステップ3201において、ユーザリスト115cに格納された、該ユーザの希望する、追加配送テキスト数を読み込む。

【0267】次にステップ3202において、個人適合度判定情報管理エリア2003から、過去所定の期間内に該ユーザに配送されなかったテキストのうち、適合度が高いものからステップ3201において読み込んだ数だけのテキストに関する、適合度判定情報を読み込む。

【0268】次にステップ3203において、ステップ3202において読み込んだテキスト識別子に対応するテキストの内容を、テキスト保存エリア2004から読

み込み、ステップ 3202 で読み込んだ適合度判定情報とともに、ワークエリア 118 に出力する。

【0269】次にステップ 3204 において、ワークエリア 118 に格納されている内容を、メールプログラム 113 が配送できる形に成形する。

【0270】ワークエリア 118 に格納された情報は、電子メールプログラム 113 により、該ユーザに配送される。

【0271】以上が、テキスト追加配送プログラム 2901 の処理手順である。

【0272】以下、図 36 に示したテキスト追加配送プログラム 2901 の処理の流れを、図 37 を用いて具体的に説明する。以下、本例では、テキスト追加配送プログラム 2901 は、システム制御プログラム 110b により、2.4 時間毎に起動されるものとするが、他の時間間隔で起動されるものとしてもよいし、所定の時刻に起動されるものとしてもよい。また、ユーザ毎に起動時間を設定してもよい。

【0273】テキスト追加配送プログラム 2901 の処理においては、以下に示すステップ 3201 ～ステップ 3204 をすべてのユーザについて繰り返す。

【0274】まず、ステップ 3201 において、ユーザリスト 115c から、所定の期間毎の追加配送希望テキスト数を取得する。この図に示した例では、ユーザ 107 は 2.4 時間毎に、2 件のテキストの追加配送を希望しているものとする。

【0275】次に、ステップ 3202 において、個人適合度判定情報管理エリア 2003 から、該ユーザに配送していないテキストのうち、上位追加配送希望数分だけのテキスト識別子を読み込む。本図に示した例では、過去 2.4 時間に、ユーザ 107 に配送されなかったテキスト 2、テキスト 4、テキスト 5 のうち、適合度の上位 2 件、即ちテキスト 2 とテキスト 5 のテキスト識別子、適合度が読み込まれる。このとき、本ステップの後に配送閾値修正提案提示ステップを追加して、該ユーザに配送したテキストの数を計数し、所定の数より少ない場合には、該ユーザに配送閾値を下げるように所定の警告文を、また所定の数より多い場合には、該ユーザに配送閾値を上げるように所定の警告文をワークエリア 118 に出力しユーザに送付することもできる。

【0276】次に、ステップ 3203 において、ステップ 3202 で読み込んだテキスト識別子に対応するテキスト情報を、テキスト保存エリア 2004 から読み込む。そして、ステップ 3202 において取得した適合度判定情報とともにワークエリア 118 に出力する。本図に示した例では、テキスト 2 とテキスト 5 の受信時刻、適合度、内容、がワークエリア 118 に出力される。

【0277】次に、ステップ 3204 において、ワークエリア 118 に格納されている情報を電子メールプログラム 113 が配送できる形に成形する。

【0278】以上が、テキスト追加配送プログラム 2901 の具体的な処理の流れである。

【0279】以上が、第五の実施例における、各プログラムの処理手順である。

【0280】以上示したように本実施例によれば、ニュース配信元から受信した全テキストや、各ユーザに対して行なった適合度算出の履歴を保存しておき、適合度が配送閾値を超えなかったテキストも、適合度が上位のものから追加配送する形態をとる。これにより、各ユーザが設定した配送閾値を超える適合度のテキストの数が、各ユーザの所望する数より少ない場合にも、所定のテキスト数を追加配送することが可能になる。したがって、ユーザにテキストが全く配送されなかった場合に、所望のテキストが出現しなかったためのか、配送閾値を高く設定しすぎたためのか判断できない、といった問題が解決できる。

【0281】ここで、本実施例では、ユーザが設定した配送閾値を超える適合度のテキストに加えて、ユーザが設定した数のテキストを配信するものとしたが、配送閾値を超えるテキストと、追加配信するテキスト数の和が、ユーザが設定した数を満たすようにする方法を用いてもよい。

【0282】次に、本発明の第六の実施例について説明する。

【0283】第四の実施例においては、配送閾値を好適な値に変更することが可能になるが、配送閾値の初期値を適切に設定することが困難である。例えば、検索条件を新たに設定したとき、その検索条件に対して、ユーザの所望するテキストにどの程度の適合度が算出されるのかがわからない。

【0284】また、第四、第五の実施例においては、時々刻々変化するテキストの配信状況に対して常に適切な配送閾値を設定するためには、ユーザが常に配送閾値の修正を行う必要があるという問題がある。

【0285】以上の問題を解決するために、本発明を適用した文書検索配送システムの第六の実施例では、ユーザが希望する配送テキスト数を設定し、その設定をもとにシステムが適時、配送閾値の設定を行うものである。

【0286】次に、本発明の第六の実施例のシステム構成を図 38 に示す。

【0287】本実施例は、図 1 に示す第一の実施例とほぼ同様の構成をとり、図 38 に示されるように希望配送数設定プログラム 3400、配送閾値自動設定プログラム 3401、および保存テキスト検索プログラム 3402 が追加される。

【0288】さらに、第四の実施例におけるものと同一の、個人適合度判定情報管理エリア 2003 とテキスト保存エリア 2004 が追加される。

【0289】また、テキスト検索プログラム 131 におけるテキスト配送判定プログラム 152a は、第四の実

施例におけるものと同一である。

【0290】本実施例では、個人適合度判定情報管理エリア2003およびテキスト保存エリア2004に格納された情報および、希望配送数設定プログラム3400によりユーザが設定した希望配送数に基づき、配送閾値自動設定プログラム3401が、各ユーザに適切な配送閾値を設定する。ここで設定された配送閾値は、テキスト検索配送制御プログラム112aにより、受信したテキストを各ユーザに配送するか否かを判定する際に用いられる。

【0291】また、検索条件を新規に登録した場合など、個人適合度判定情報が保存されていないユーザに対しては、保存テキスト検索プログラム3402により、テキスト保存エリア2004に保存されているテキストに対して適合度算出を行い、その結果を個人適合度判定情報管理エリア2003に格納する。

【0292】以下に、第六の実施例における、各プログラムの処理手順について説明する。

【0293】まず、第六の実施例におけるシステム制御プログラム110cの処理手順を図39のPAD図を用いて説明する。

【0294】本実施例におけるシステム制御プログラム110cの処理手順は、第一の実施例におけるシステム制御プログラム110とほぼ同じであるが、ステップ3500～ステップ3504の処理が加わる。

【0295】ステップ303の処理につづき、ステップ3502において、希望配送数設定プログラム3400を起動する。

【0296】次に、ステップ3503において、保存テキスト検索プログラム3402を起動する。

【0297】次に、ステップ3504において、配送閾値自動設定プログラム3401を起動する。

【0298】一方、ステップ300の繰り返し処理において、ステップ301、ステップ302の処理の後、ステップ3500で、ユーザ107から、所定の時間毎に希望する配送テキスト数の設定要求が送られてきているかを調べる。ここで、該設定要求が送られてきている場合には、ステップ3505で希望配送数設定プログラム3400を起動する。

【0299】次に、ステップ3501で、現在の時刻が所定の条件を満たす時刻であるかを判定する。ここで、所定の条件を満たす時刻である場合には、ステップ3506において配送閾値自動設定プログラム3401を起動する。例えば、「毎日午前零時に配送閾値自動設定プログラム3401を起動する」という条件を設定して、配送閾値自動設定プログラム3401を起動することができる。

【0300】以上が、システム制御プログラム110cの処理手順である。

【0301】以下、図39に示したステップ3502ま

たはステップ3505でシステム制御プログラムにより起動される、希望配送数設定プログラム3400の処理手順を、図40のPAD図を用いて説明する。

【0302】希望配送数設定プログラムでは、まずステップ3600において、ユーザ107から指定時間と、その指定時間あたりに配送を希望するテキスト数を取得する。

【0303】次に、ステップ3601において、該ユーザのユーザ識別子をもとに、ユーザリスト115dにおける該ユーザの指定時間および希望配送テキスト数を、ステップ3600で取得したそれぞれの値で更新する。ここで、ユーザリスト115d中の各ユーザの指定時間および希望配送テキスト数の初期設定値は、管理者が予め定めた値であってもよいし、ユーザ107が検索条件に登録するときに入力するものとしてもよい。

【0304】以上が、希望配送数設定プログラム3400の処理手順である。

【0305】以下、図39に示したステップ3503でシステム制御プログラム110cにより起動される、保存テキスト検索プログラム3402の処理手順を、図41のPAD図を用いて説明する。

【0306】本プログラムは、検索条件を新規に登録した直後など、適合度算出の履歴がない場合に、保存されているテキストに対して適合度を算出するものである。

【0307】保存テキスト検索プログラム3402は、ステップ3700において、所定期間内にテキスト保存エリア2004に保存された全てのテキストについて、ステップ3701およびステップ3702を繰り返し実行する。

【0308】まずステップ3701において、テキスト検索走査プログラム150および適合度算出プログラム151を起動し、検索条件登録制御プログラム111により登録された検索条件に対する、テキストの適合度を算出する。

【0309】次にステップ3702において、ステップ3701で算出された適合度を、適合度判定情報管理エリア2003に格納する。

【0310】以上が、保存テキスト検索プログラム3402の処理手順である。以下、図39に示したステップ3504またはステップ3506でシステム制御プログラム110cにより起動される、配送閾値自動設定プログラム3401の処理手順を、図42のPAD図を用いて説明する。

【0311】本プログラムは、過去に受信したテキストに対して算出した適合度の分布をもとに、各ユーザ毎に適切な配送閾値を設定するものである。

【0312】配送閾値自動設定プログラム3401は、ステップ3800において、ユーザリスト115dに格納された全てのユーザについて、ステップ3801～ステップ3804を繰り返し実行する。

【0313】まずステップ3801において、ユーザリスト115dに格納された指定時間と、指定時間あたりに該ユーザが希望する配送テキスト数を取得する。

【0314】次に、ステップ3802において、該ユーザのユーザ識別子をもとに、個人適合度判定情報管理エリア2003から、過去所定の期間内にニュース配信元106から受信したテキストに関する、該ユーザの個人適合度判定情報を読み込む。

【0315】次に、ステップ3803において、ステップ3802で読み込んだ個人適合度判定情報を用いて、所定の算出式を用いて、新たに設定する配送閾値を算出する。

【0316】次に、ステップ3804において、ステップ3803で算出した配送閾値を、ユーザリスト115dの該ユーザの配送閾値格納部分に設定する。

【0317】以上が、配送閾値自動設定プログラム3401の処理手順である。

【0318】以下、図42に示した配送閾値自動設定プログラム3401の処理の流れを、図43を用いて具体的に説明する。

【0319】以下、本例では、配送閾値自動設定プログラム3401は、システム制御プログラム110cにより、24時間毎に起動されるものとするが、他の時間間隔で起動されるものとしてもよいし、所定の時刻に起動するものとしてもよい。また、ユーザ毎に起動時刻を設定してもよい。

【0320】配送閾値自動設定プログラム3401の処理においては、以下に示すステップ3801～ステップ3804をすべてのユーザについて繰り返す。

【0321】まず、ステップ3801において、ユーザリスト115dから、指定時間と希望配送テキスト数を取得する。本図に示した例では、ユーザ1が48時間に4件の頻度でテキスト配送を希望しており、指定時間として「48時間」、希望配送テキスト数として「4件」、という情報が取得される。

【0322】次に、ステップ3802において、個人適合度判定情報管理エリア2003から、ユーザ107に対する、過去指定期間内の個人適合度判定情報における適合度を読み込む。本図に示した例では、過去24時間に配送されたテキスト1、テキスト2、テキスト3のそれぞれの適合度である「10.0」、「5.0」、「7.0」が読み込まれる。

【0323】次に、ステップ3803において、ステップ3802で読み込んだ適合度をもとに、所定の算出方法を用いて、新しい配送閾値を算出する。本図に示した例では、まず48時間あたり4件という条件を、24時間あたり2件という条件に換算する。ついで、ステップ3802で読み込んだ適合度のうち、上位2件目（「7.0」）とその次の適合度（「5.0」）の平均値「6.0」が、新たな配送閾値として算出される。本ステップ38

03における配送閾値の算出方法としては、他の方法を用いてもよい。

【0324】次に、ステップ3804において、ユーザリスト115dにおける該当ユーザの配送閾値として、ステップ3803で算出した配送閾値を格納する。

【0325】以上が、配送閾値自動設定プログラム3401の具体的な処理の流れである。

【0326】以上が、第六の実施例における、各プログラムの処理手順である。

【0327】以上示したように本実施例によれば、各ユーザに対して行なった適合度算出の履歴を保存しておき、その情報を参照することで、適切な配送閾値を自動的に修正する形態をとる。これにより、常に各ユーザにとって適切な配送閾値を設定しておくことが可能となり、ユーザは逐一配送閾値を設定し直すという手間を省くことができる。したがって、時々刻々変化するテキストの配信状況に対して常に適切な配送閾値を設定するためには、ユーザが常に配送閾値の修正を行う必要がある、といった問題を解決できる。

【0328】また、検索条件を新たに設定した場合にも、過去のテキストを走査し、その検索条件に対する適合度を算出する形態をとるため、適切な配送閾値を算出し設定することが可能になる。即ち、新たに設定した検索条件に対して、ユーザの所望するテキストにどの程度の適合度が算出されるのかがわからない、といった問題を解決できる。

【0329】ここで、第四、第五の実施例における配送閾値設定支援情報提示プログラム2001、配送閾値更新プログラム2000、および配送閾値設定試行プログラム2002と、第六の実施例における配送閾値自動設定プログラム、希望配送数設定プログラム、および保存テキスト検索プログラム3402とをひとつのシステムに共に実装することも可能である。この場合、ユーザまたはシステム管理者が、どちらの実施例によるシステムを使用するかを選択し登録する、配送条件設定モード選択プログラムを追加することにより、配送閾値を設定するか、あるいは配信テキスト数を設定するかをシステム使用者が選択することも可能となる。

【0330】また、本実施例において説明した配送テキストの選択方法は、上述した有限オートマトンや拡張BM法といったテキストの検索方法に依るものではなく、ここで説明したもの以外の検索方法を用いたシステムでも同様に用いることができるものである。

【0331】また、本実施例において説明した配送テキストの選択方法は、検索条件に対するテキストの適合度算出方法によるものではなく、第一、第二、第三の実施例に示した適合度算出方法、またそれ以外の適合度算出方法でも同様に用いることができるものである。

【0332】なお、以上第一から第六の実施例において、ディスプレイ100、キーボード101、中央演算

処理装置（CPU）102、主メモリ104およびこれらと結ぶバス103から構成される文書検索配送システムは、図1、図21、図33、図38におけるニュース配信元106と通信回線105の中間や、通信回線105とユーザ107の中間のいずれのネットワーク上の位置に配置されても構わない。

【0333】

【発明の効果】 本発明を利用することで、複数ユーザの検索条件に対する適合度を算出し、ユーザ毎に配送閾値との比較を行うため、重要な情報を要求するユーザに対しては適合度の高いテキストを配送することが可能となる。

【図面の簡単な説明】

【図1】 本発明の第一の実施例の構成を示す図である。

【図2】 従来技術1の概要を説明する図である。

【図3】 本発明の第一の実施例におけるシステム制御プログラム110の処理手順を示すPAD図である。

【図4】 本発明の第一の実施例における検索条件登録制御プログラム111の処理手順を示すPAD図である。

【図5】 本発明の第一の実施例におけるテキスト検索配送制御プログラム112の処理手順を示すPAD図である。

【図6】 本発明の第一の実施例における検索用オートマトン作成プログラム123の処理手順を示すPAD図である。

【図7】 本発明の第一の実施例におけるテキスト検索プログラム131の処理手順を示すPAD図である。

【図8】 本発明の第一の実施例における検索条件登録処理の概要を説明する図である。

【図9】 本発明の第一の実施例におけるテキスト検索処理の概要を説明する図である。

【図10】 本発明の第一の実施例における検索条件登録制御プログラム111の具体的な処理の流れを説明する図である。

【図11】 本発明の第一の実施例におけるテキスト検索プログラム131の具体的な処理の流れを説明する図である。

【図12】 本発明の第二の実施例における有限オートマトン114とユーザリスト115の接続方法を説明する図である。

【図13】 本発明の第二の実施例におけるテキスト検索プログラム131aの処理手順を示すPAD図である。

【図14】 本発明の第二の実施例におけるテキスト検索プログラム131aの具体的な処理手順を説明する図である。

【図15】 本発明の第三の実施例における検索条件登録制御プログラム111aの構成を示す図である。

【図16】 本発明の第三の実施例における検索条件登録制御プログラム111aの具体的な処理手順を説明する図である。

【図17】 本発明の第三の実施例における検索用オートマトン作成プログラム123aの処理手順を示すPAD図である。

【図18】 本発明の第三の実施例における検索条件登録制御プログラム111aの具体的な処理手順を説明する図である。

【図19】 本発明の第三の実施例におけるテキスト検索プログラム151aの処理手順を示すPAD図である。

【図20】 従来技術2における特徴文字列抽出処理の流れを説明する図である。

【図21】 本発明の第四の実施例の構成を示す図である。

【図22】 本発明の第四の実施例におけるシステム制御プログラム110aの処理手順を示すPAD図である。

【図23】 本発明の第四の実施例における配送閾値更新プログラム2000の処理手順を示すPAD図である。

【図24】 本発明の第四の実施例における配送閾値設定支援情報提示プログラム2001の処理手順を示すPAD図である。

【図25】 本発明の第四の実施例における配送閾値設定支援情報提示プログラム2001の具体的な処理の流れを説明する図である。

【図26】 本発明の第四の実施例における配送閾値設定支援情報提示プログラム2001の出力例を示す図である。

【図27】 本発明の第四の実施例における配送閾値設定支援情報提示プログラム2001の出力例を示す図である。

【図28】 本発明の第四の実施例における配送閾値設定試行プログラム2002の処理手順を示すPAD図である。

【図29】 本発明の第四の実施例における配送閾値設定試行プログラム2002の具体的な処理の流れを説明する図である。

【図30】 本発明の第四の実施例における配送閾値設定試行プログラム2002の出力例を示す図である。

【図31】 本発明の第四の実施例におけるテキスト検索配送制御プログラム112aの処理手順を示すPAD図である。

【図32】 本発明の第四の実施例におけるテキスト検索配送制御プログラム112aの具体的な処理の流れを説明する図である。

【図33】 本発明の第五の実施例の構成を示す図である。

【図34】 本発明の第五の実施例におけるシステム制御プログラム110bの処理手順を示すPAD図である。

【図35】 本発明の第五の実施例における追加配送テキスト数設定プログラム2900の処理手順を示すPAD図である。

【図36】 本発明の第五の実施例におけるテキスト追加

配送プログラム 2901 の処理手順を示す PAD 図である。

【図 37】本発明の第五の実施例におけるテキスト追加配送プログラム 2901 の具体的な処理の流れを説明する図である。

【図 38】本発明の第六の実施例の構成を示す図である。

【図 39】本発明の第六の実施例におけるシステム制御プログラム 110c の処理手順を示す PAD 図である。

【図 40】本発明の第六の実施例における希望配送数設定プログラム 3400 の処理手順を示す PAD 図である。

【図 41】本発明の第六の実施例における保存テキスト検索プログラム 3402 の処理手順を示す PAD 図である。

【図 42】本発明の第六の実施例における配送閾値自動設定プログラム 3401 の処理手順を示す PAD 図である。

【図 43】本発明の第六の実施例における配送閾値自動設定プログラム 3401 の具体的な処理の流れを説明する図である。

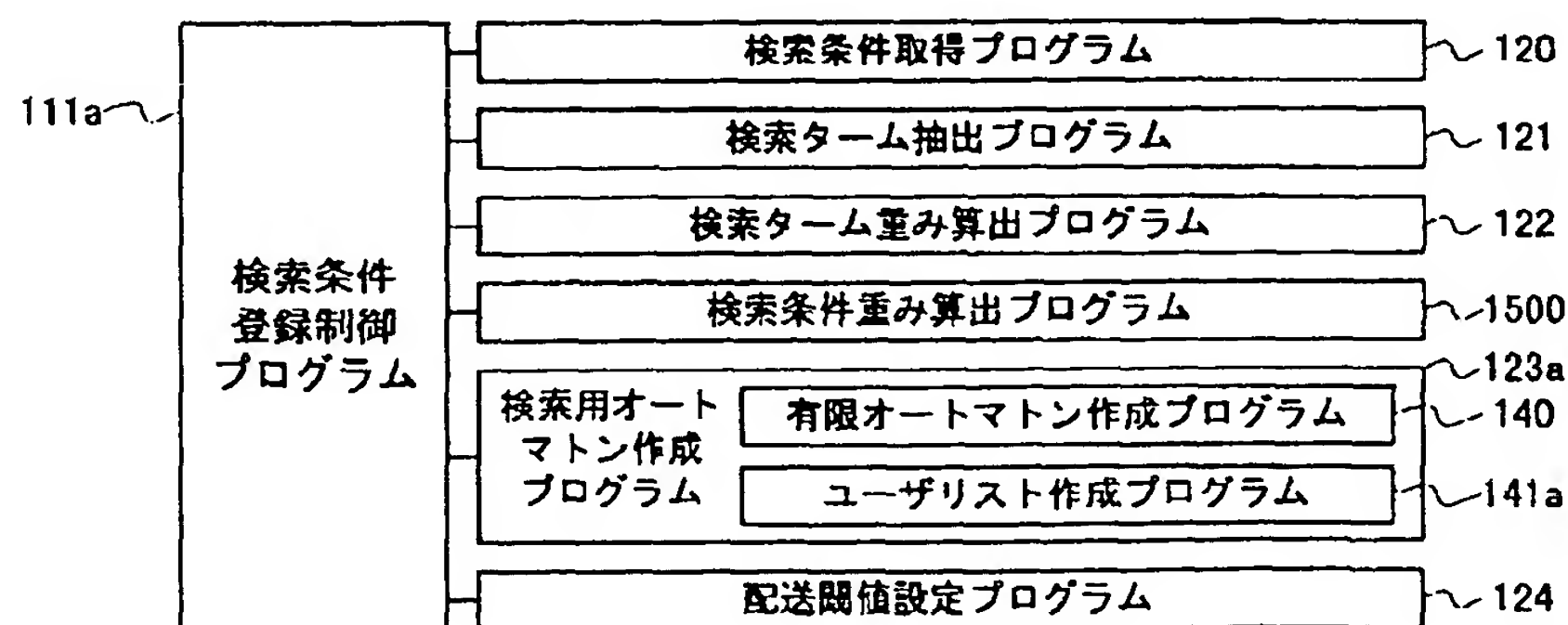
【符号の説明】

100 ディスプレイ
101 キーボード
102 中央演算処理装置 (CPU)
103 バス
104 主メモリ
105 通信回線
106 ニュース配信元
107 ユーザ
110、110a、110b、110c システム制御プログラム

111、111a 検索条件登録制御プログラム
112、112a テキスト検索配送制御プログラム
113 電子メールプログラム
114 有限オートマトン
115、115a、115b、115c、115d ユーザリスト
116 検索ターム重みテーブル
117 適合度管理テーブル
118 ワークエリア
120 検索条件取得プログラム
121 検索ターム抽出プログラム
122 検索ターム重み算出プログラム
123、123a 検索用オートマトン作成プログラム
124 配送閾値設定プログラム
130、130a テキスト取得プログラム
131、131a テキスト検索プログラム
132 テキスト成形プログラム
140 有限オートマトン作成プログラム
141 ユーザリスト作成プログラム
150 テキスト走査プログラム
151、151a 適合度算出プログラム
152 テキスト配送判定プログラム
2000 配送閾値更新プログラム
2001 配送閾値設定支援情報提示プログラム
2002 配送閾値設定試行プログラム
2003 個人適合度判定情報管理エリア
2004 テキスト保存エリア
2900 追加配送テキスト数設定プログラム
2901 テキスト追加配送プログラム
3400 希望配送数設定プログラム
3401 配送閾値自動設定プログラム
3402 保存テキスト検索プログラム

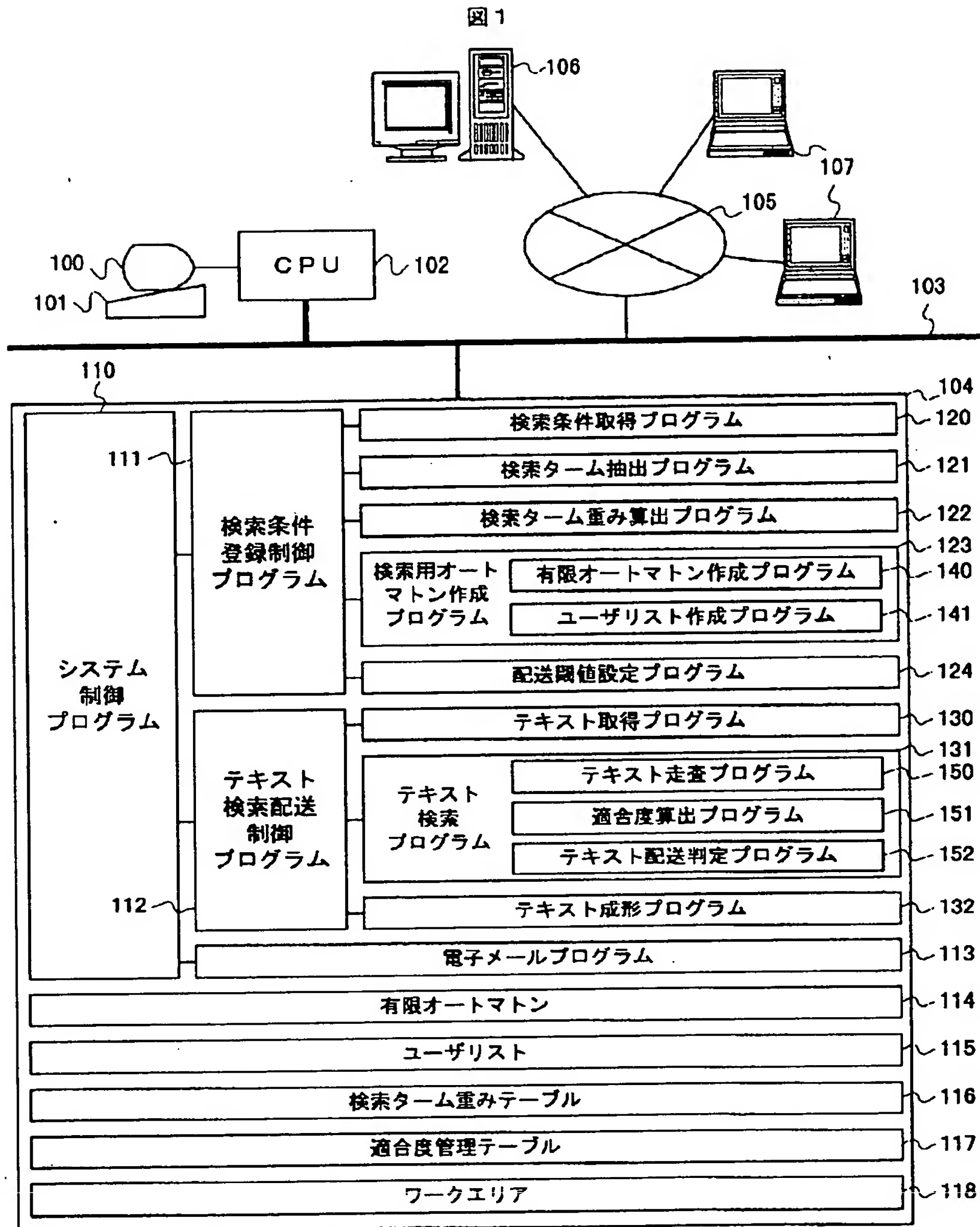
【図 15】

図 15



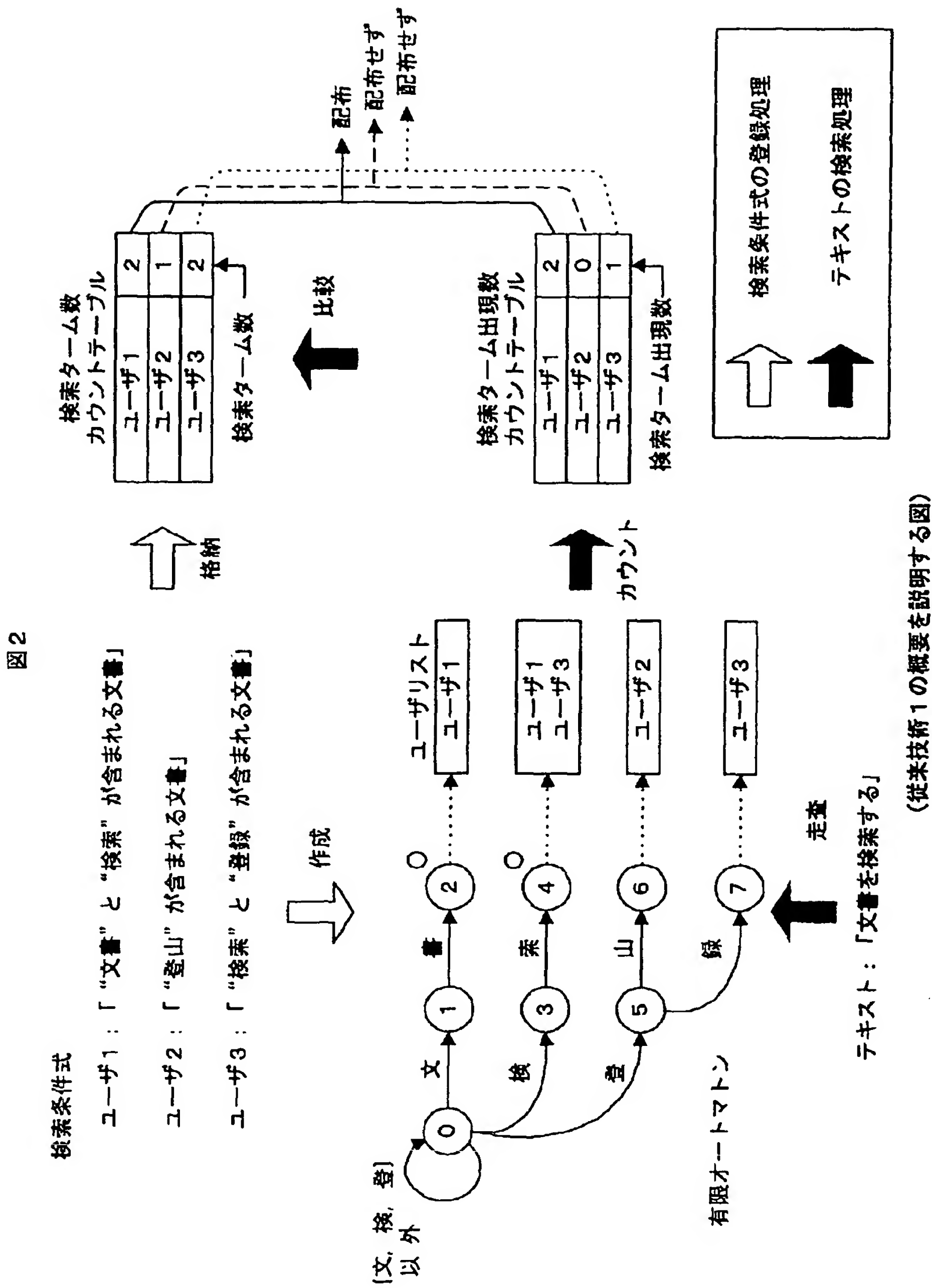
(本発明の第三の実施例における検索条件登録制御プログラム 111a の構成を示す図)

【図1】



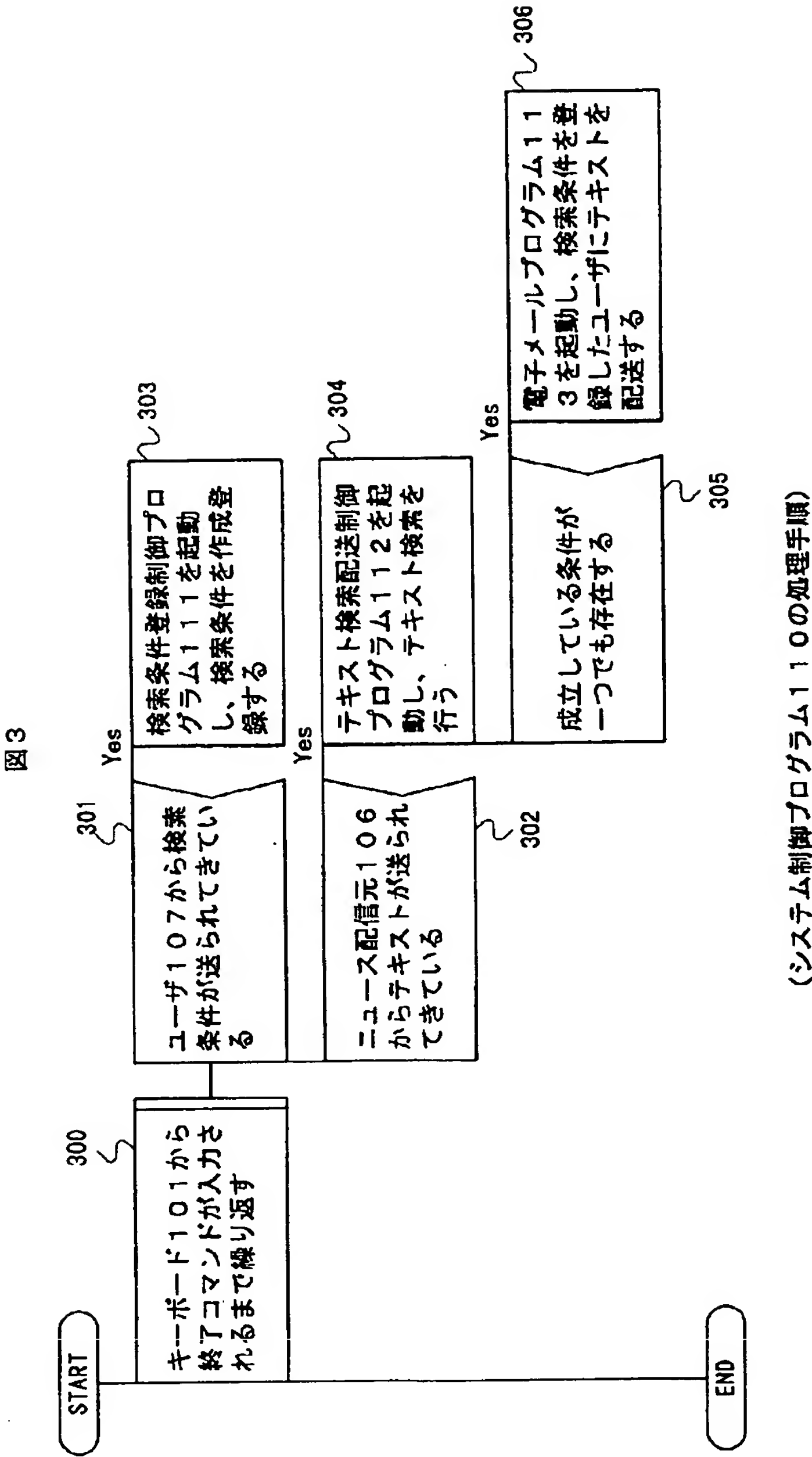
(本発明の第一の実施例の構成を示す図)

【図2】



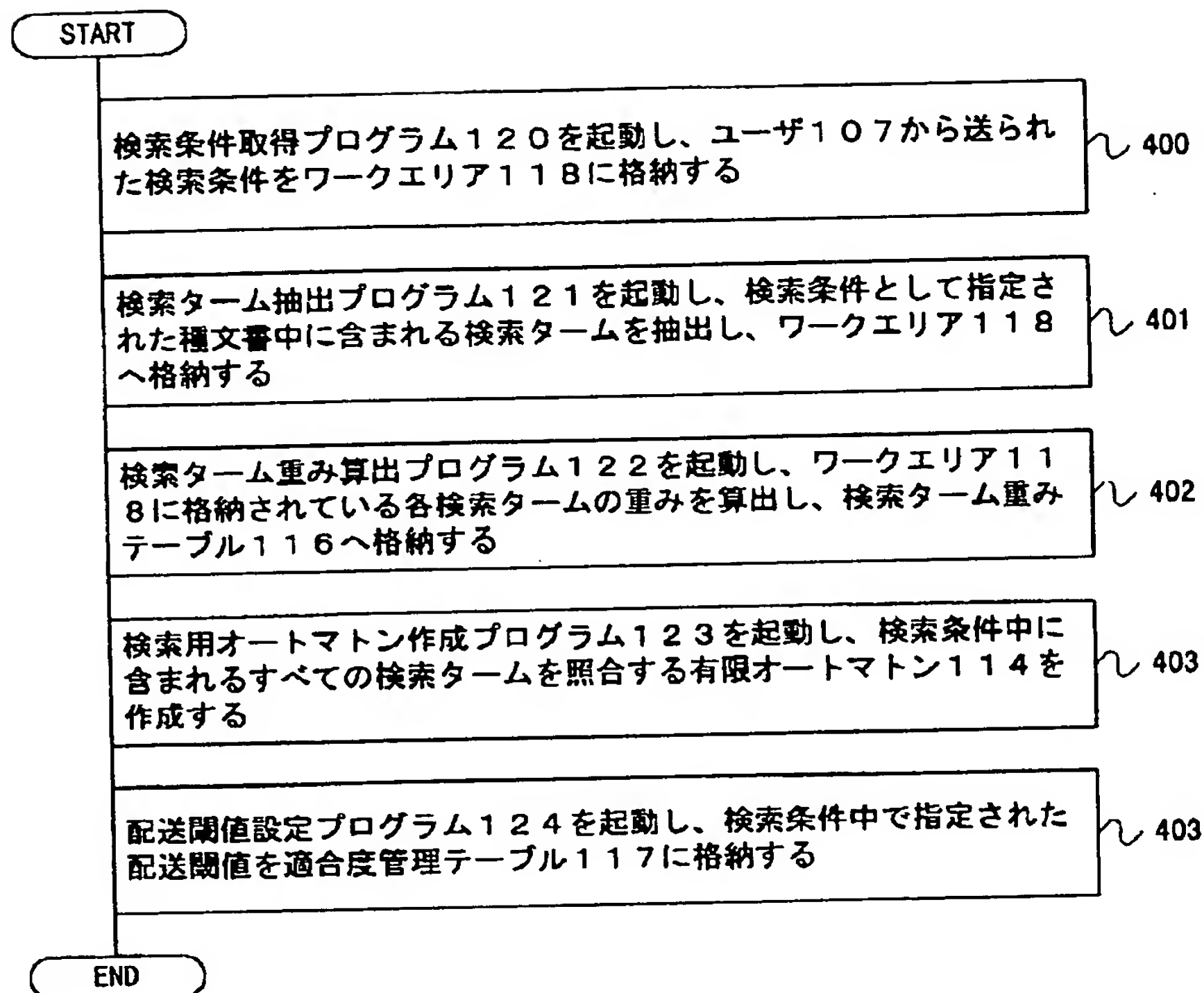
(従来技術1の概要を説明する図)

【図3】



【図4】

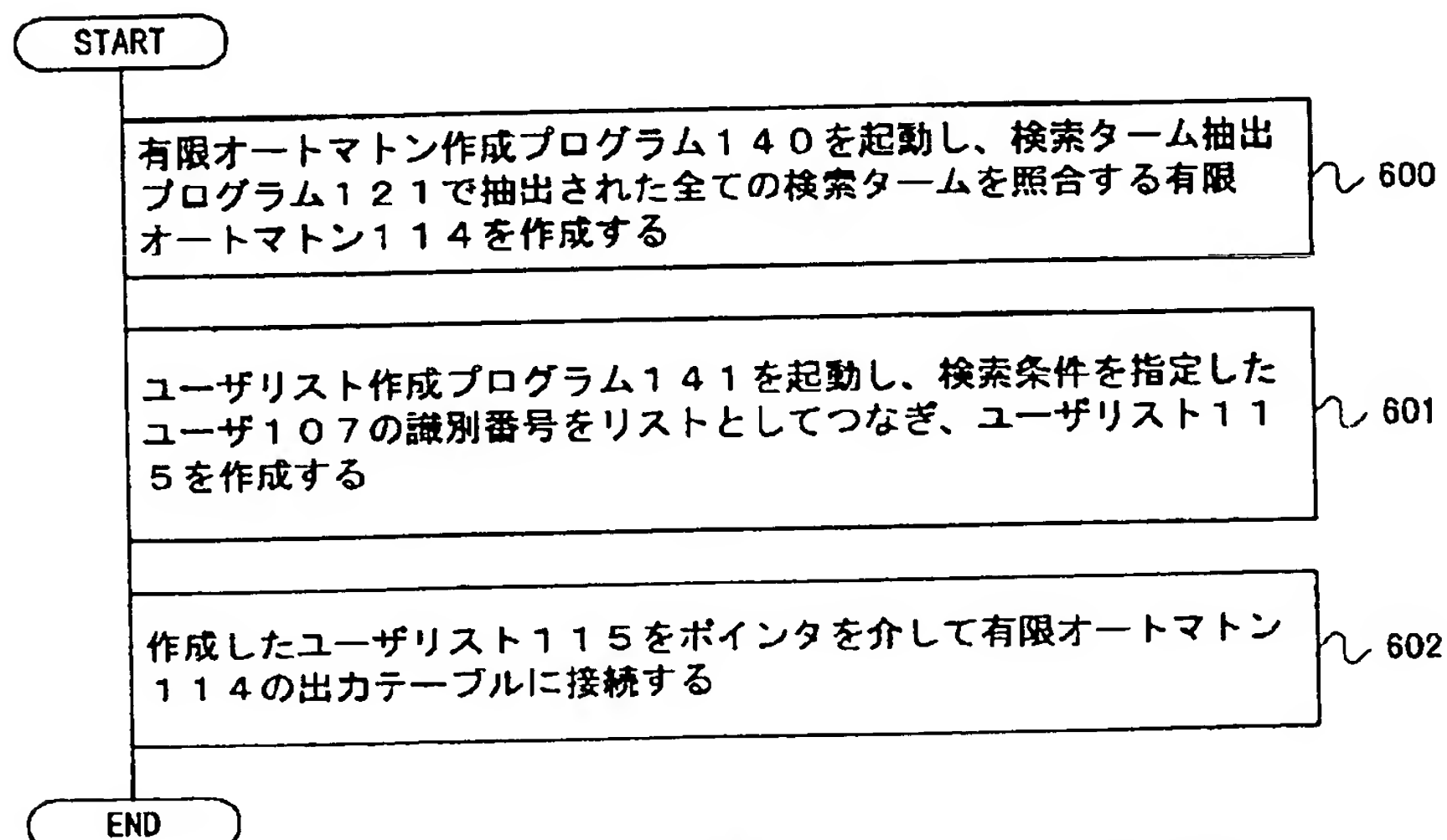
図4



(検索条件登録制御プログラム111の処理手順)

【図6】

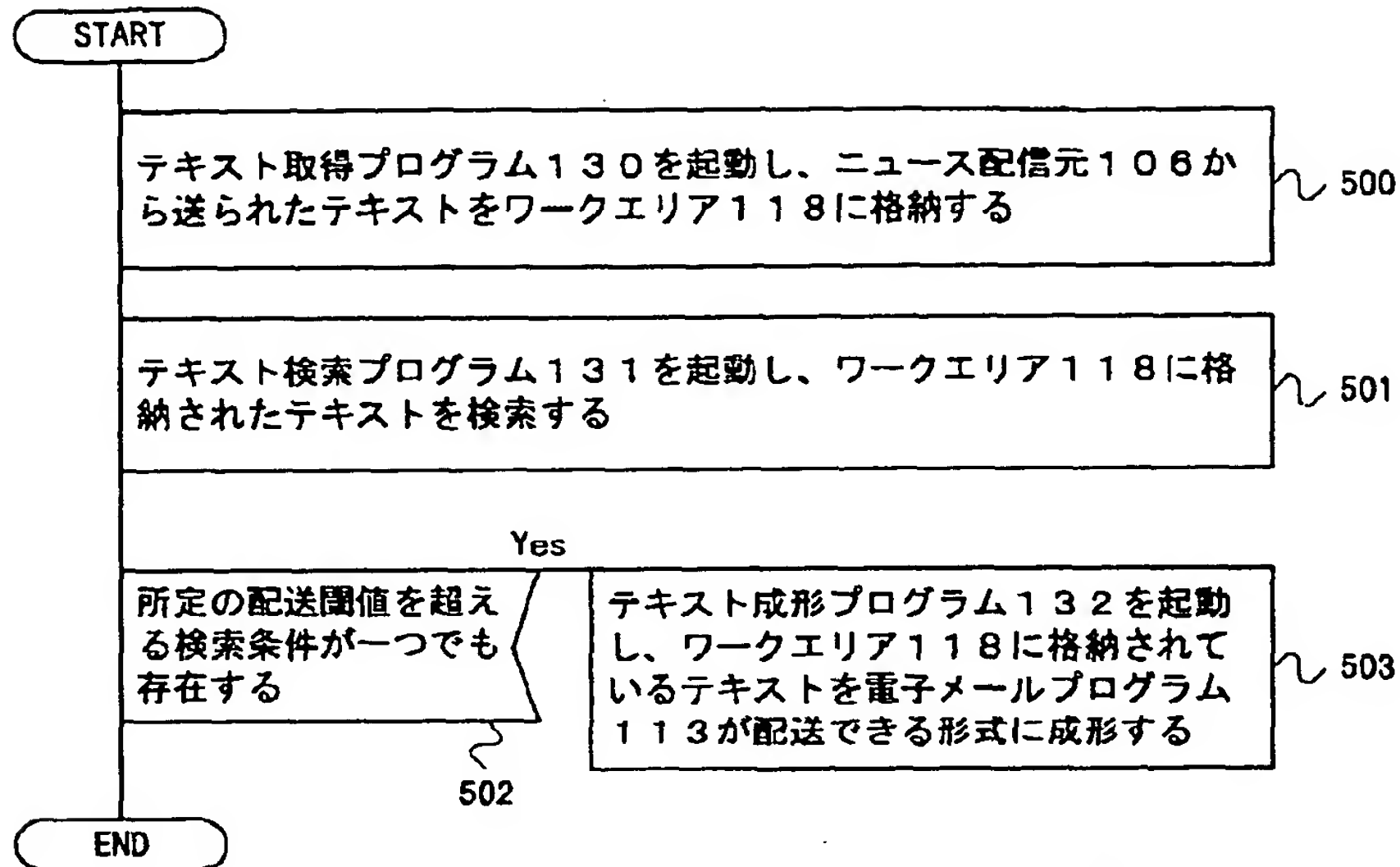
図6



(検索用オートマトン作成プログラム123の処理手順)

【図5】

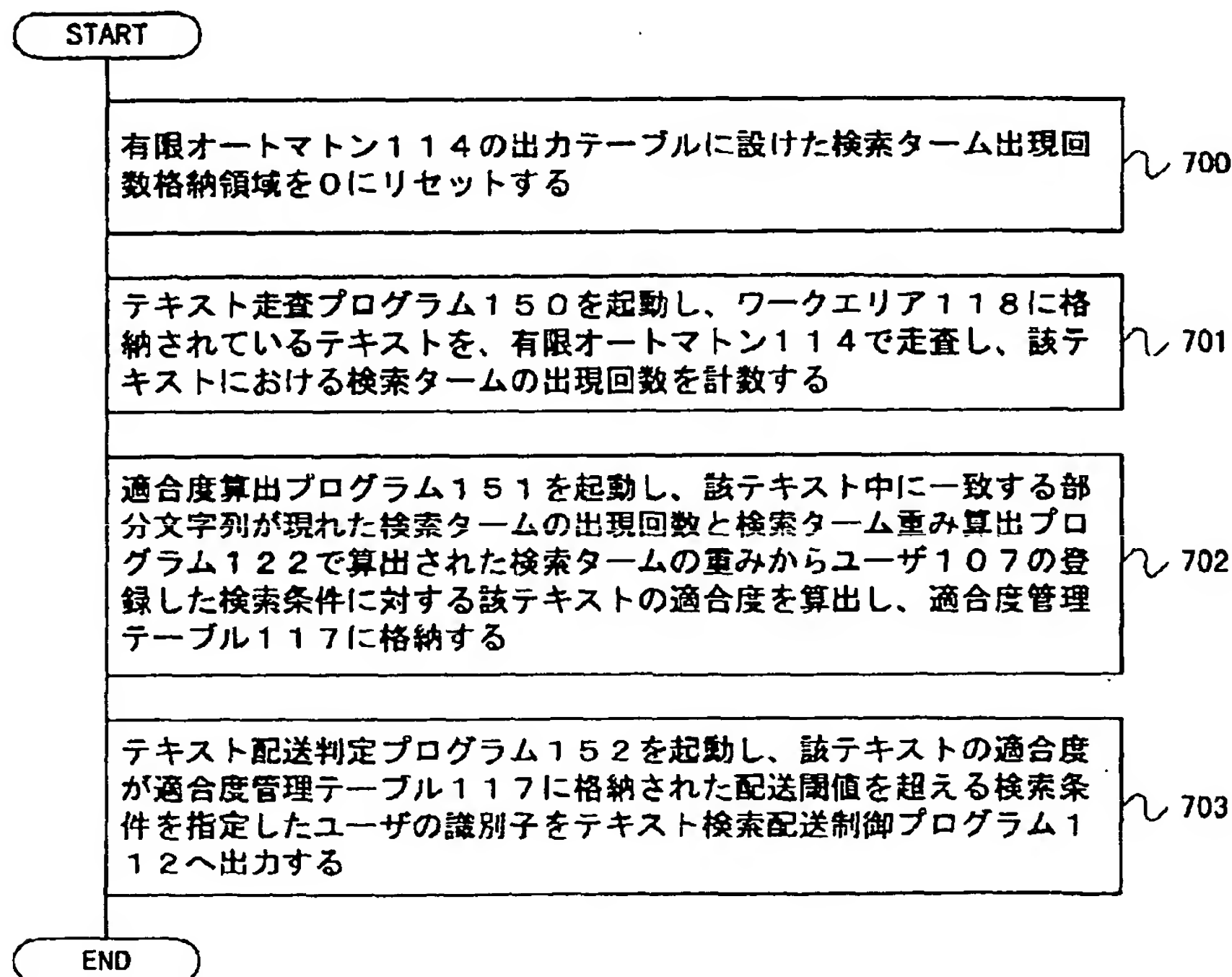
図5



(テキスト検索配送制御プログラム112の処理手順)

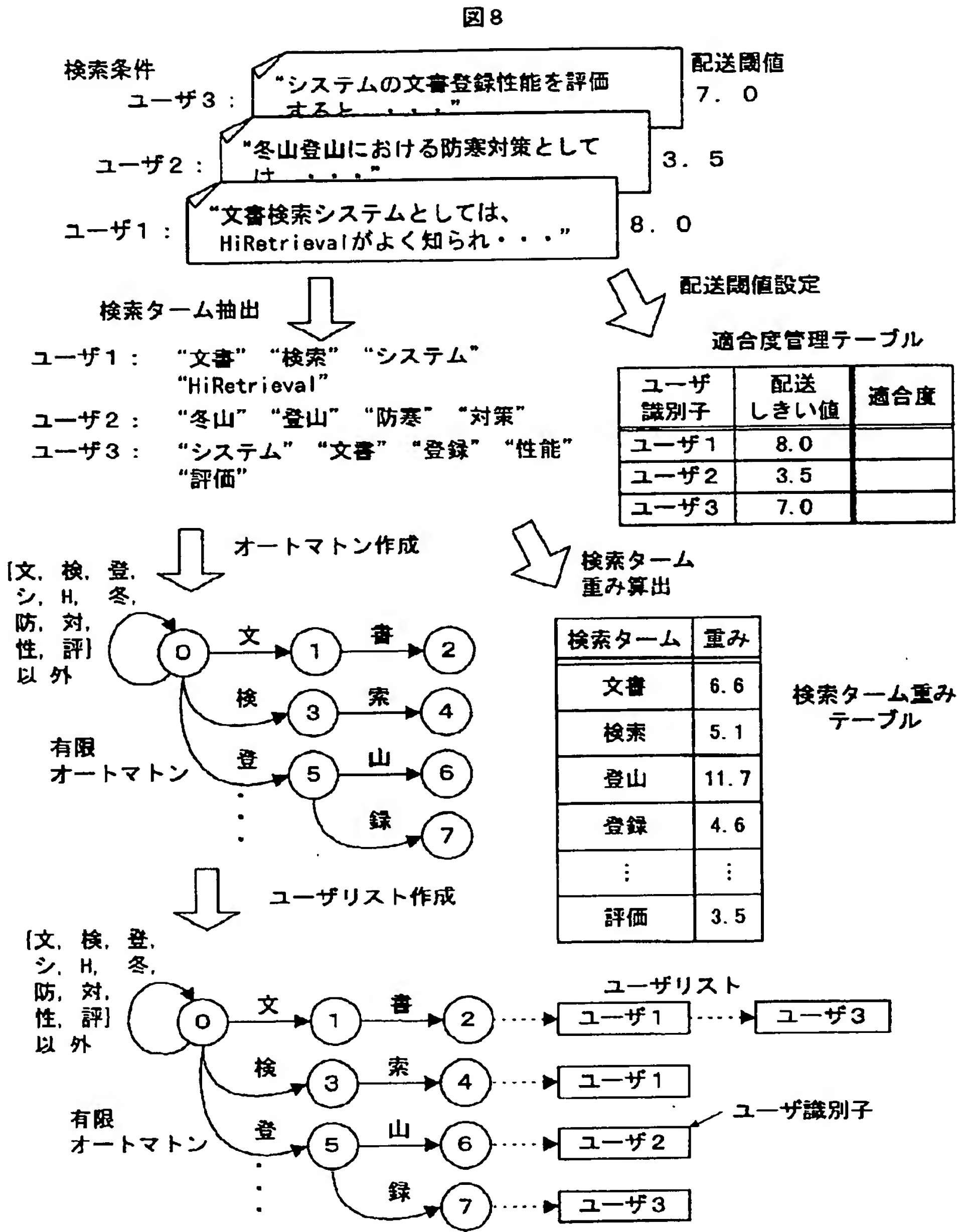
【図7】

図7



(テキスト検索プログラム131の処理手順)

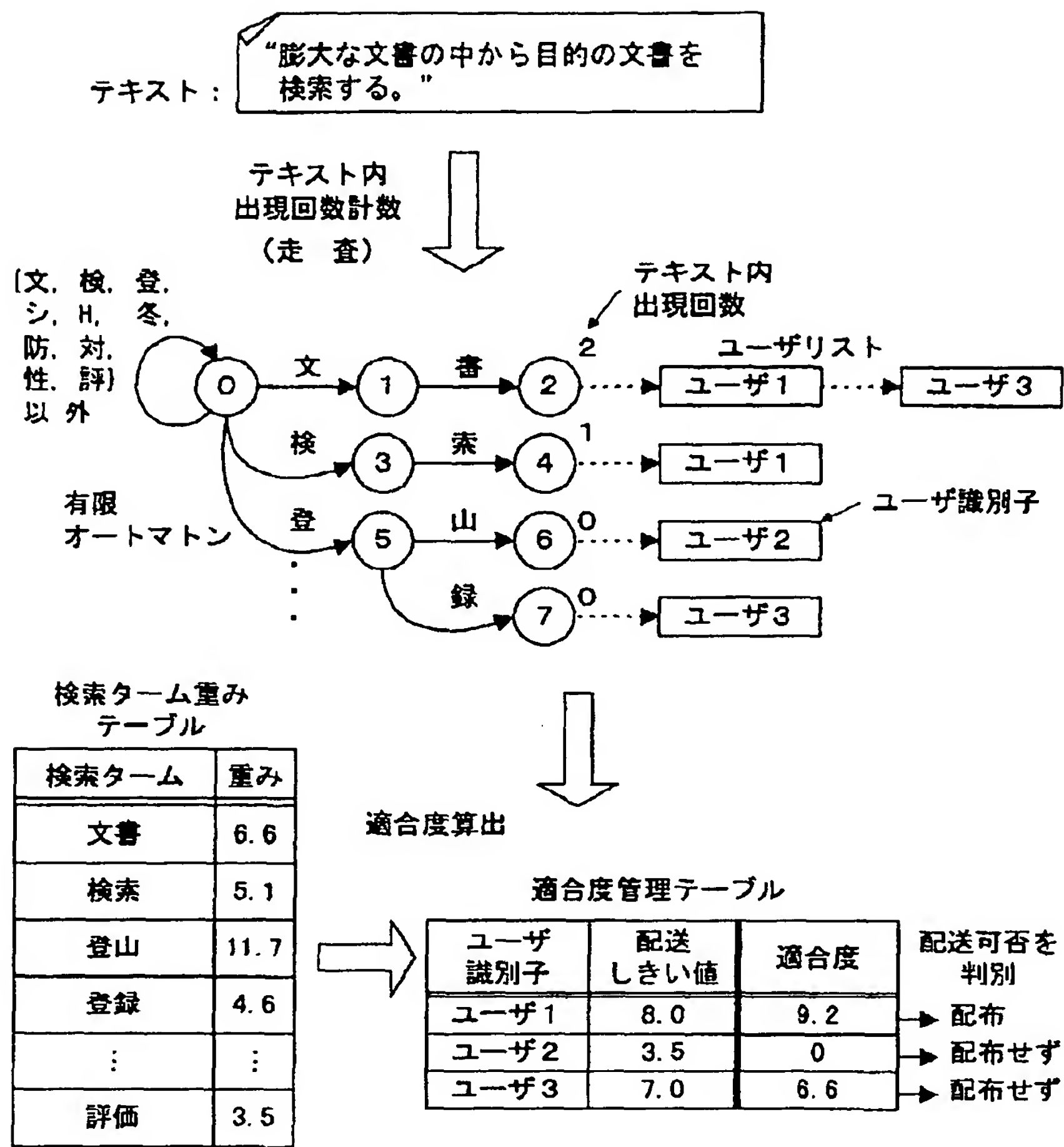
【図8】



(第一の実施例における検索条件登録処理の概要を説明する図)

【図 9】

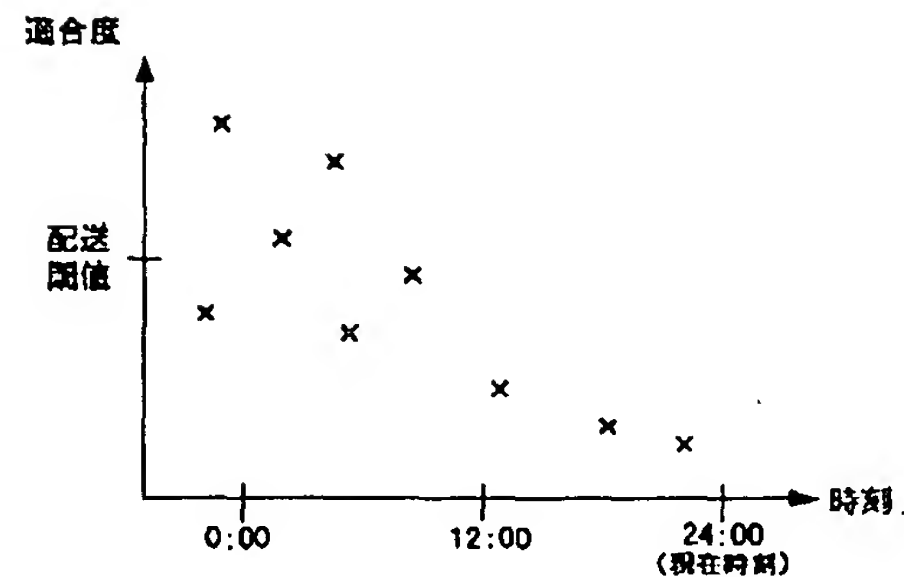
図 9



(本発明の第一の実施例におけるテキスト検索処理の概要を説明する図)

【図 26】

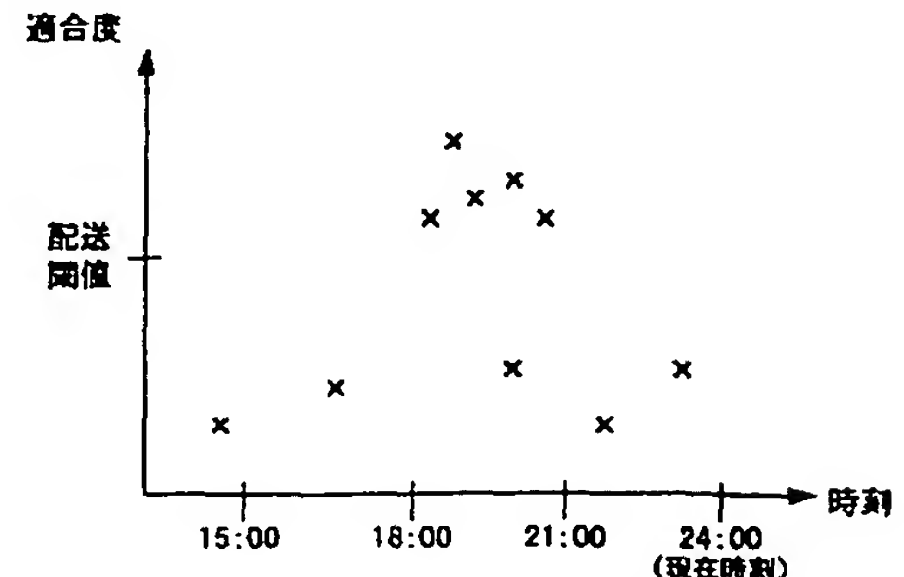
図 26



(配送閾値設定支援情報提示プログラム 2001 の出力例を説明する図)

【図 27】

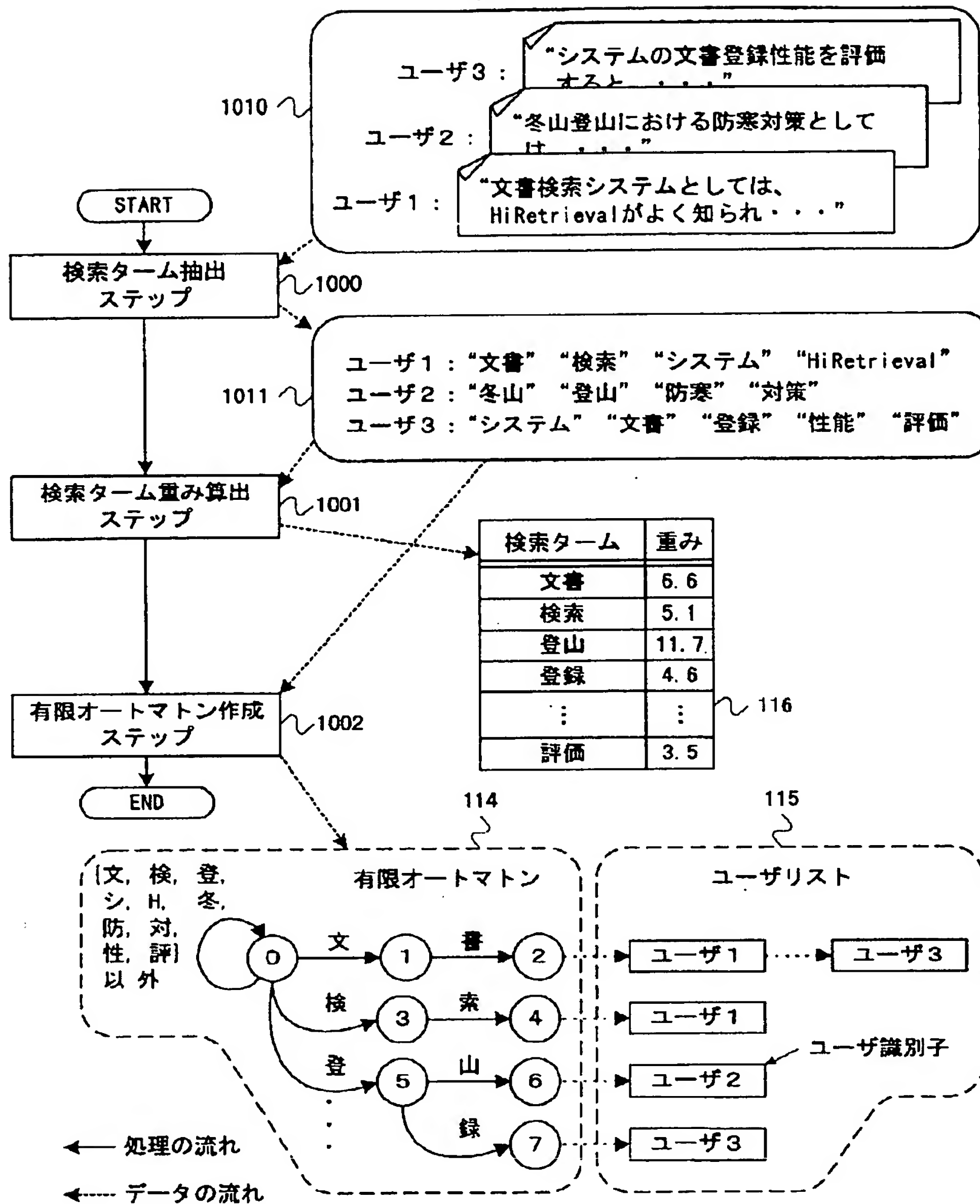
図 27



(配送閾値設定支援情報提示プログラム 2001 の出力例を説明する図)

【図10】

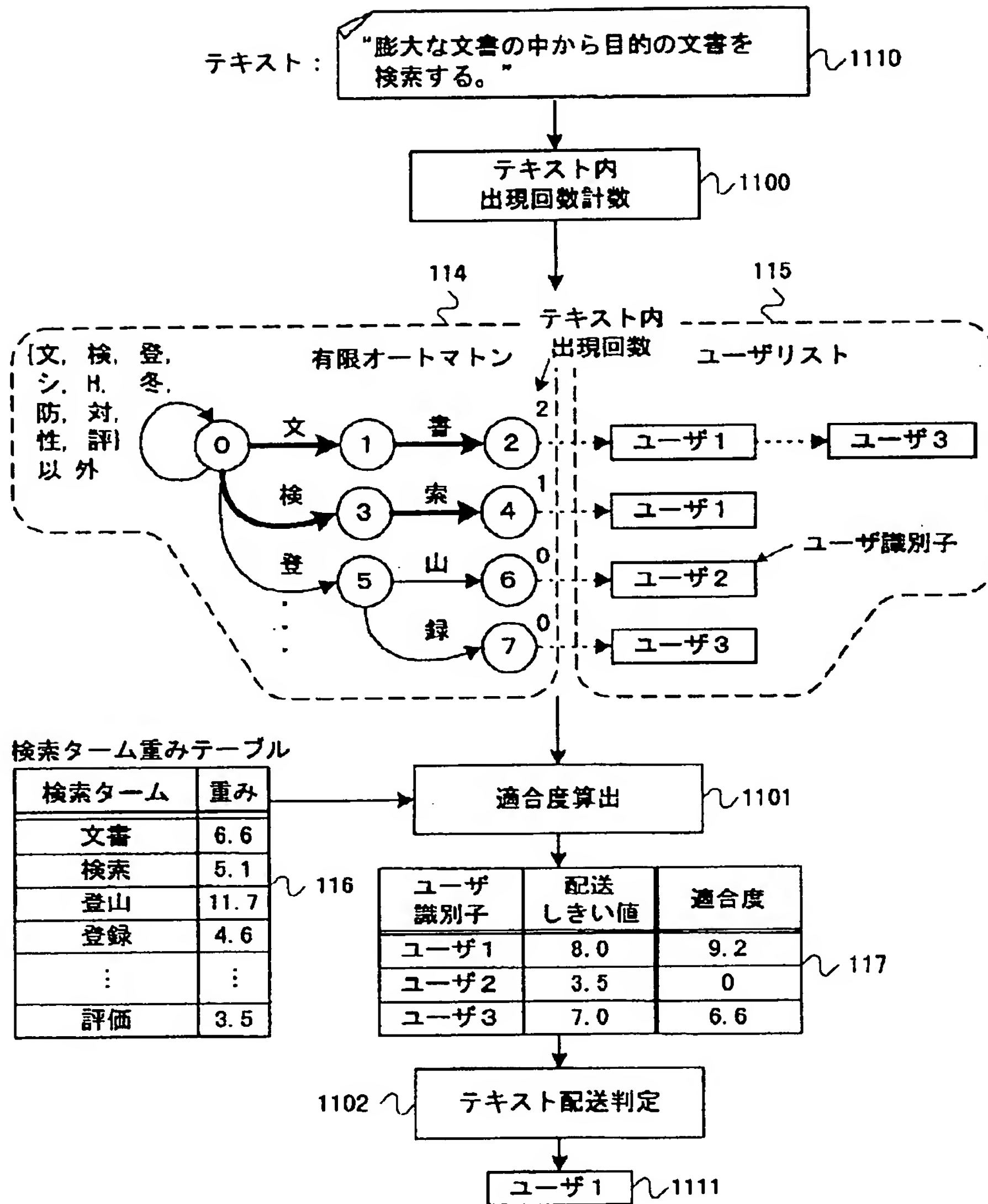
図10



(検索条件登録制御プログラム111の具体的な処理の流れを説明する図)

【図11】

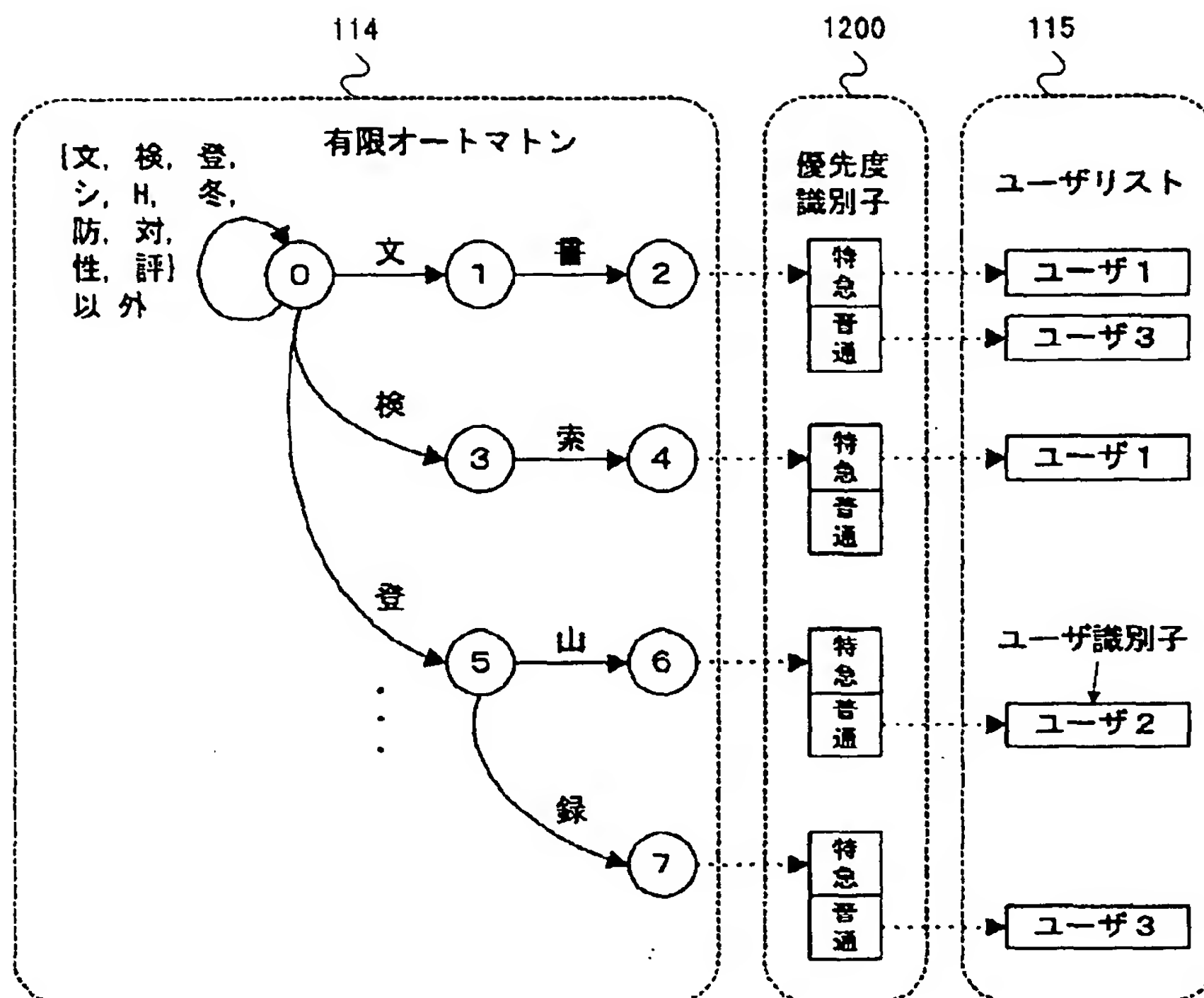
図11



(テキスト検索プログラム131の具体的な処理の流れを説明する図)

【図 12】

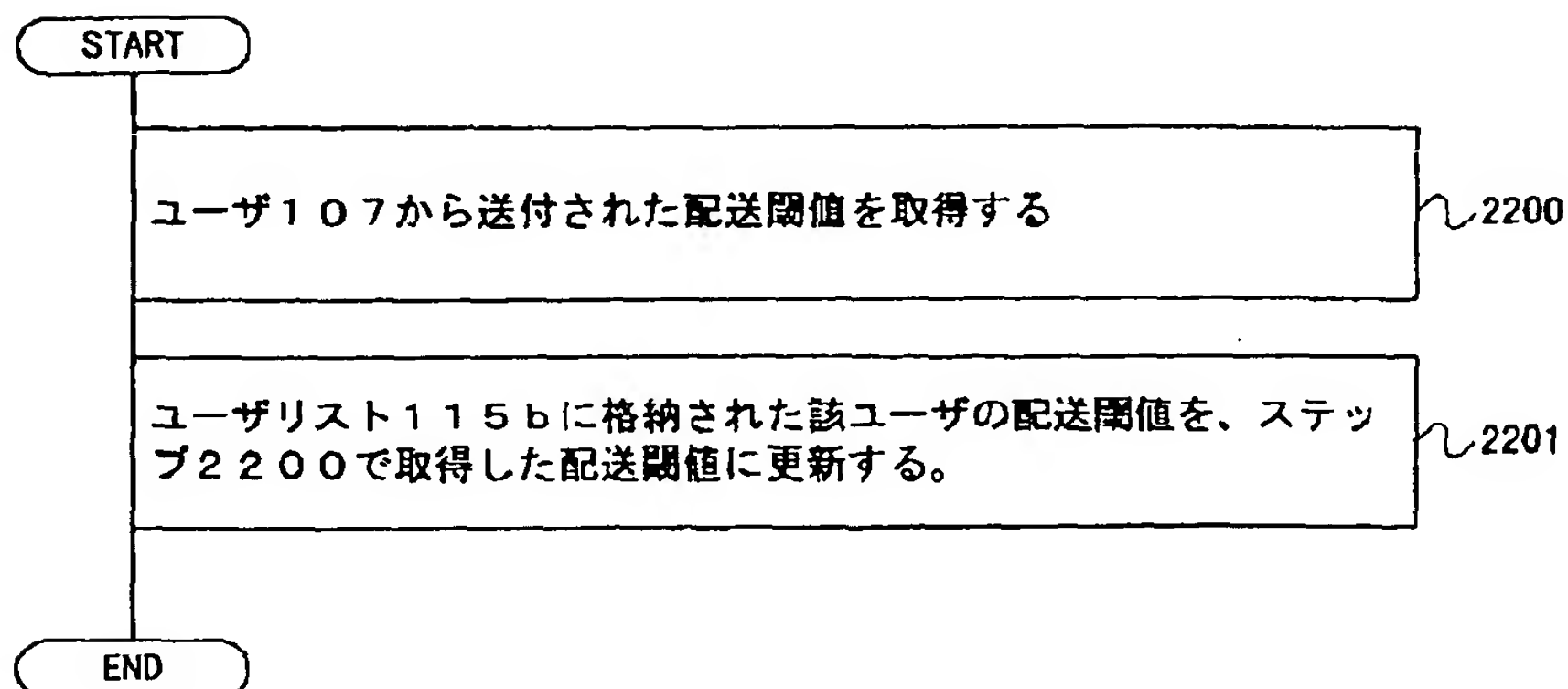
図 12



(本発明の第二の実施例におけるユーザリストの接続例)

【図 23】

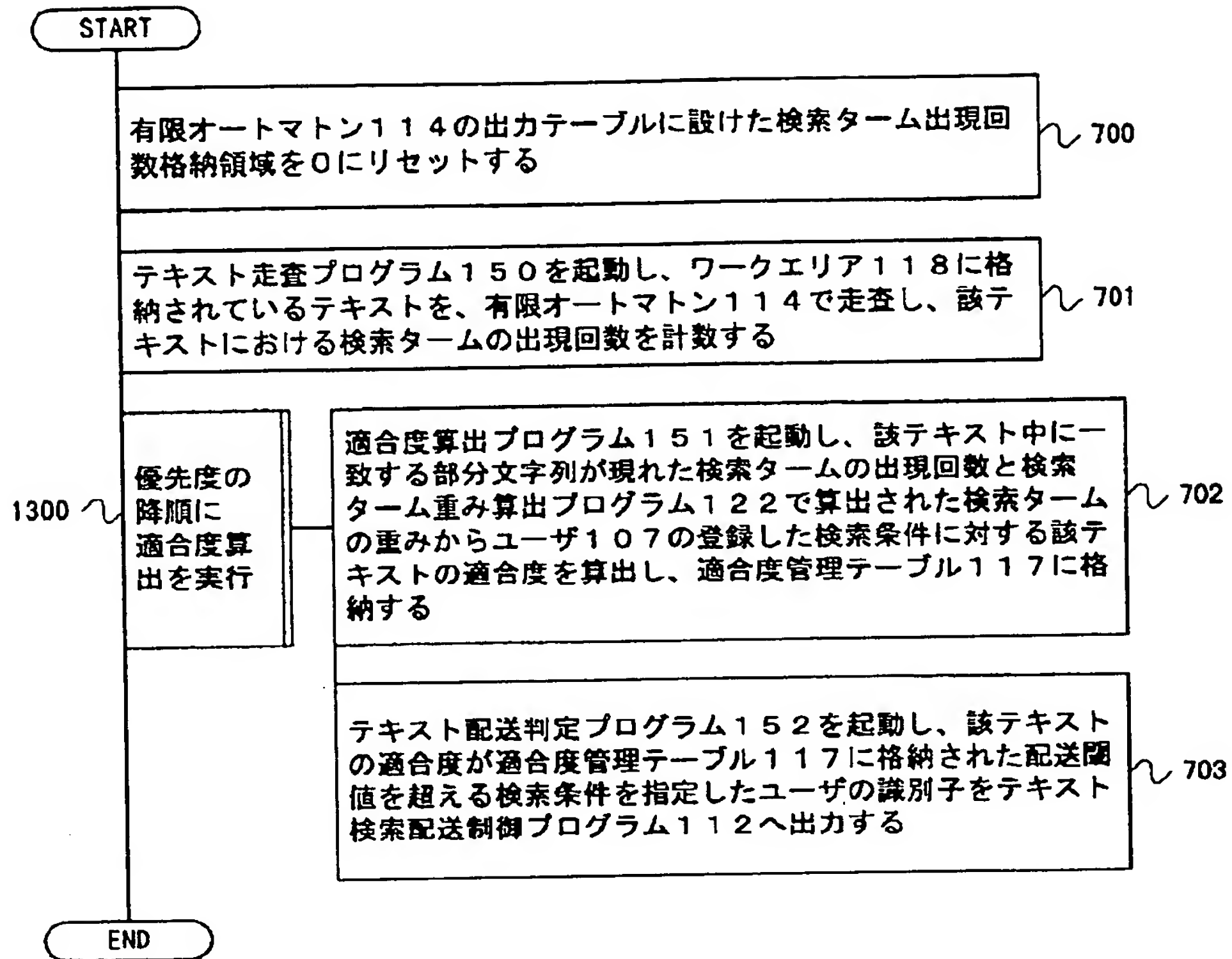
図 23



(配送閾値更新プログラム2000の処理手順)

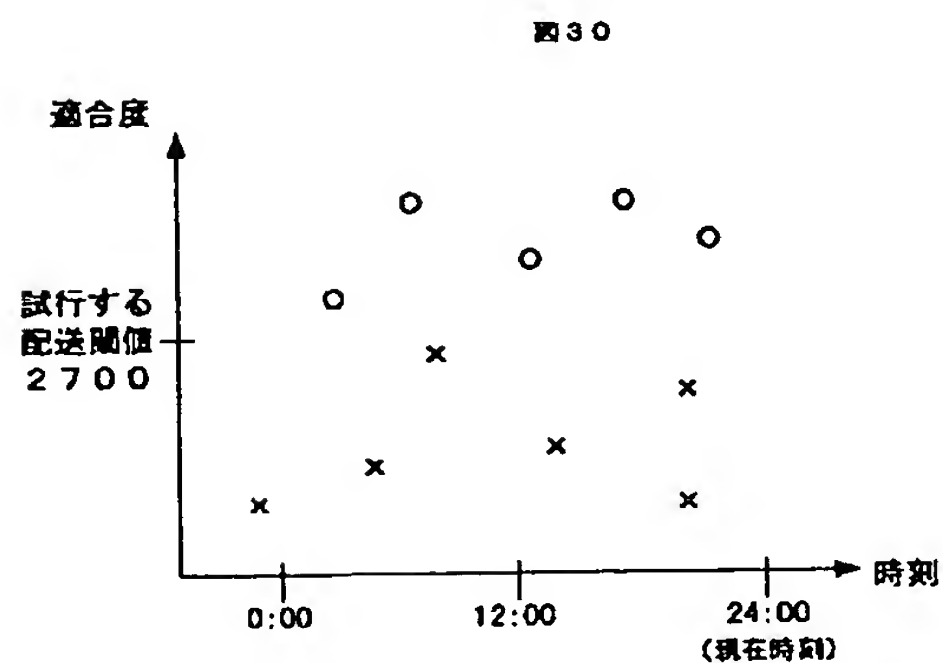
【図 13】

図 13



(テキスト検索プログラム131aの処理手順)

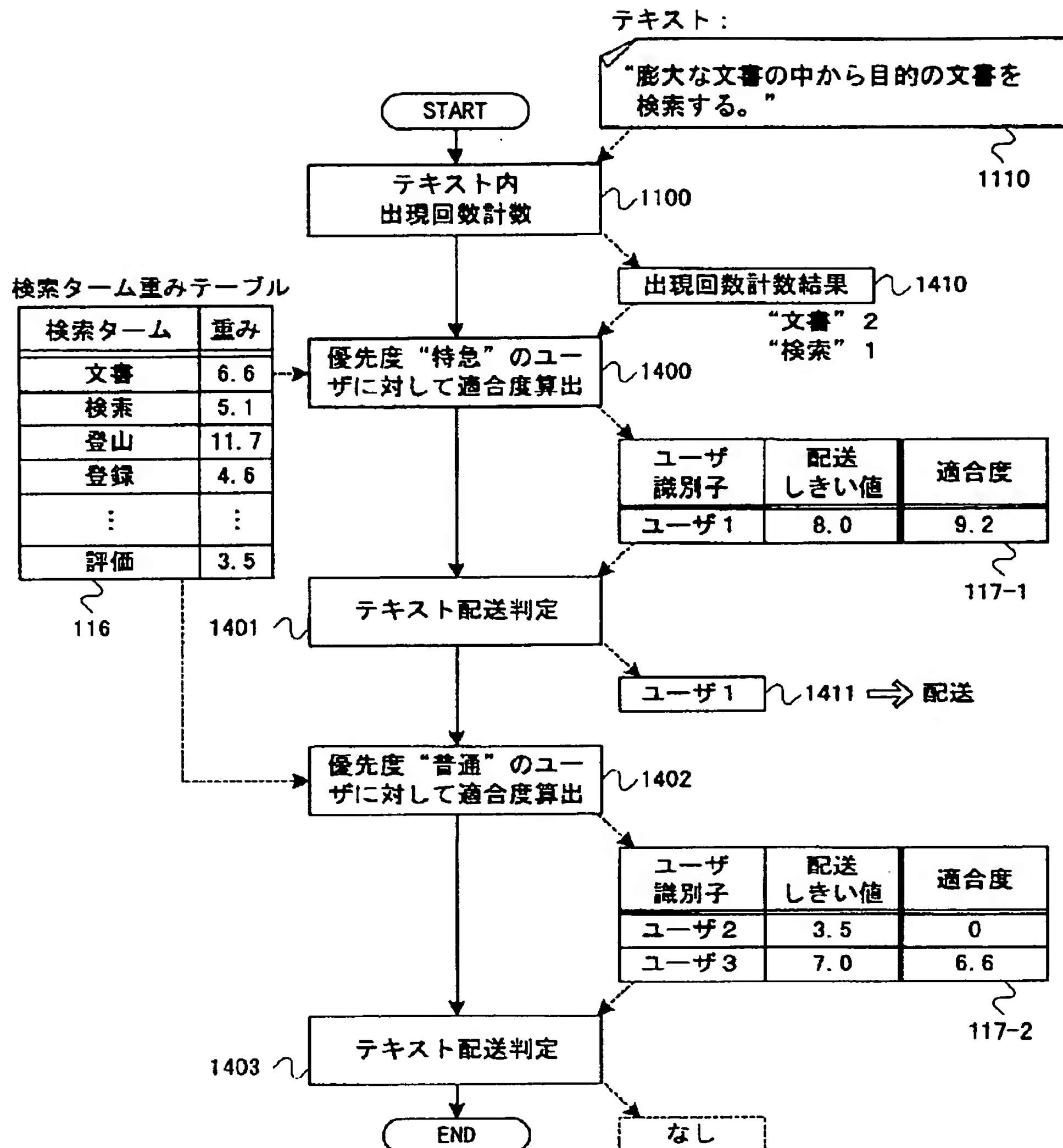
【図 30】



(配送閾値設定試行プログラム2002の出力例を説明する図)

【図 14】

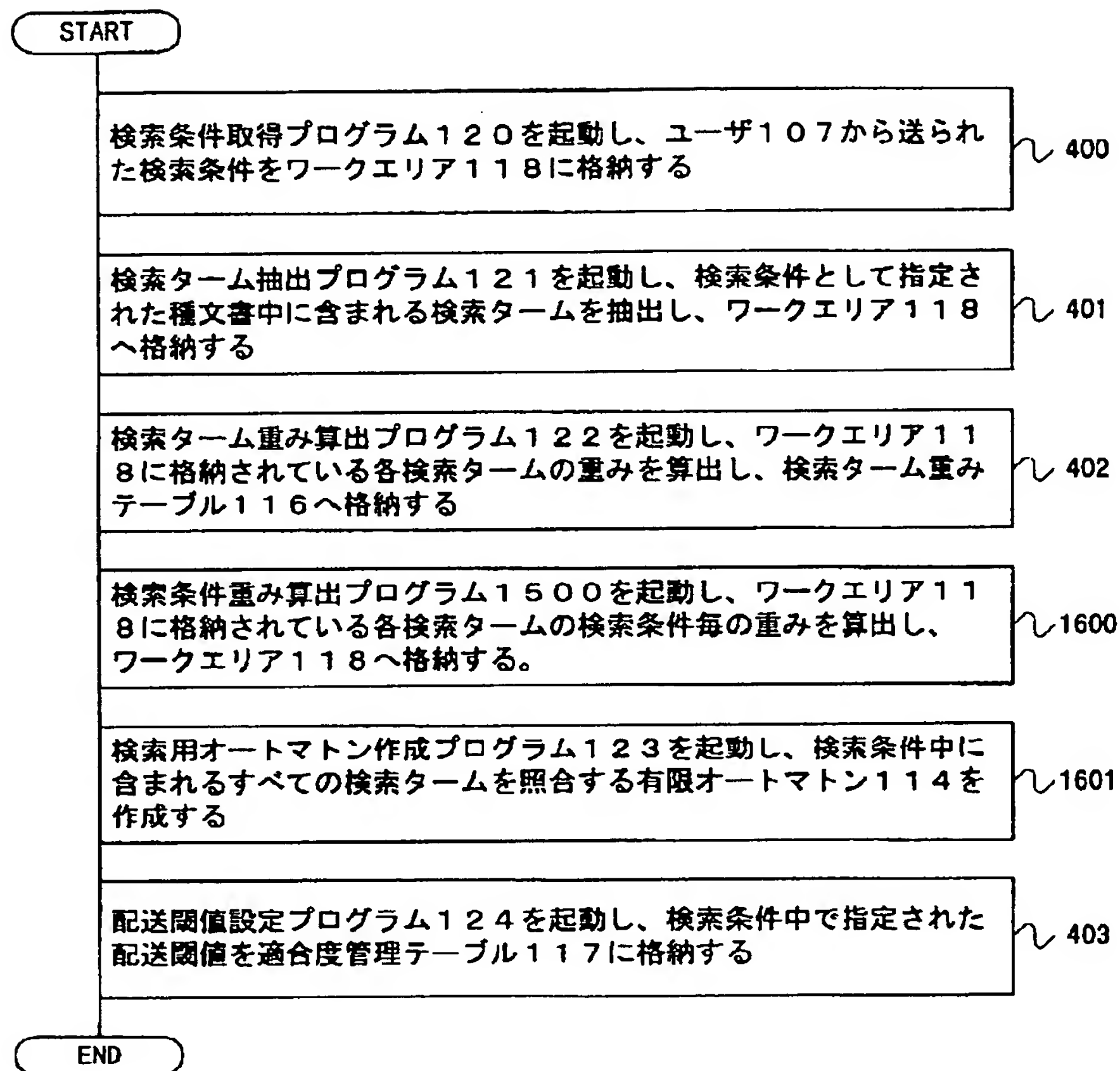
図 14



(テキスト検索プログラム 131a の具体的な処理の流れを説明する図)

【図 16】

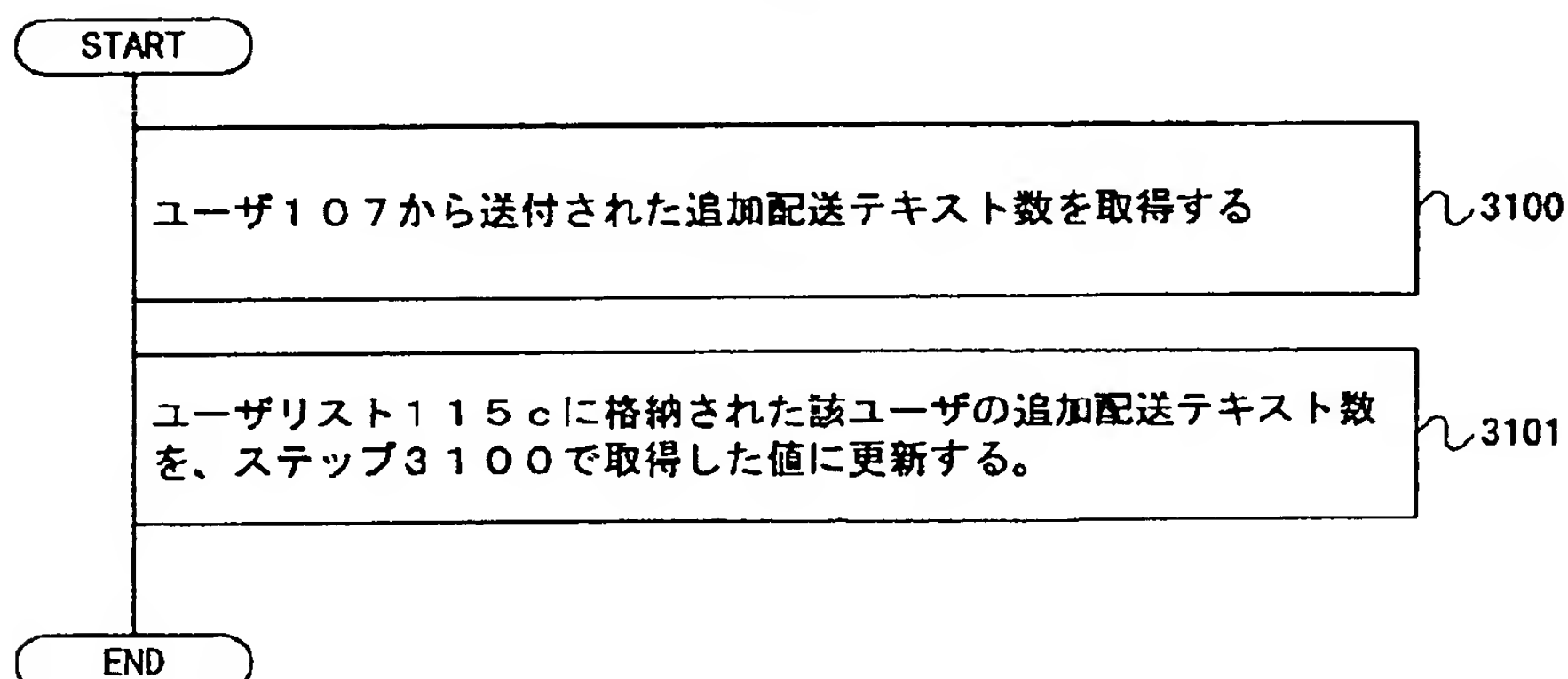
図 16



(本発明の第三の実施例における検索条件登録制御プログラム111aの処理手順)

【図 35】

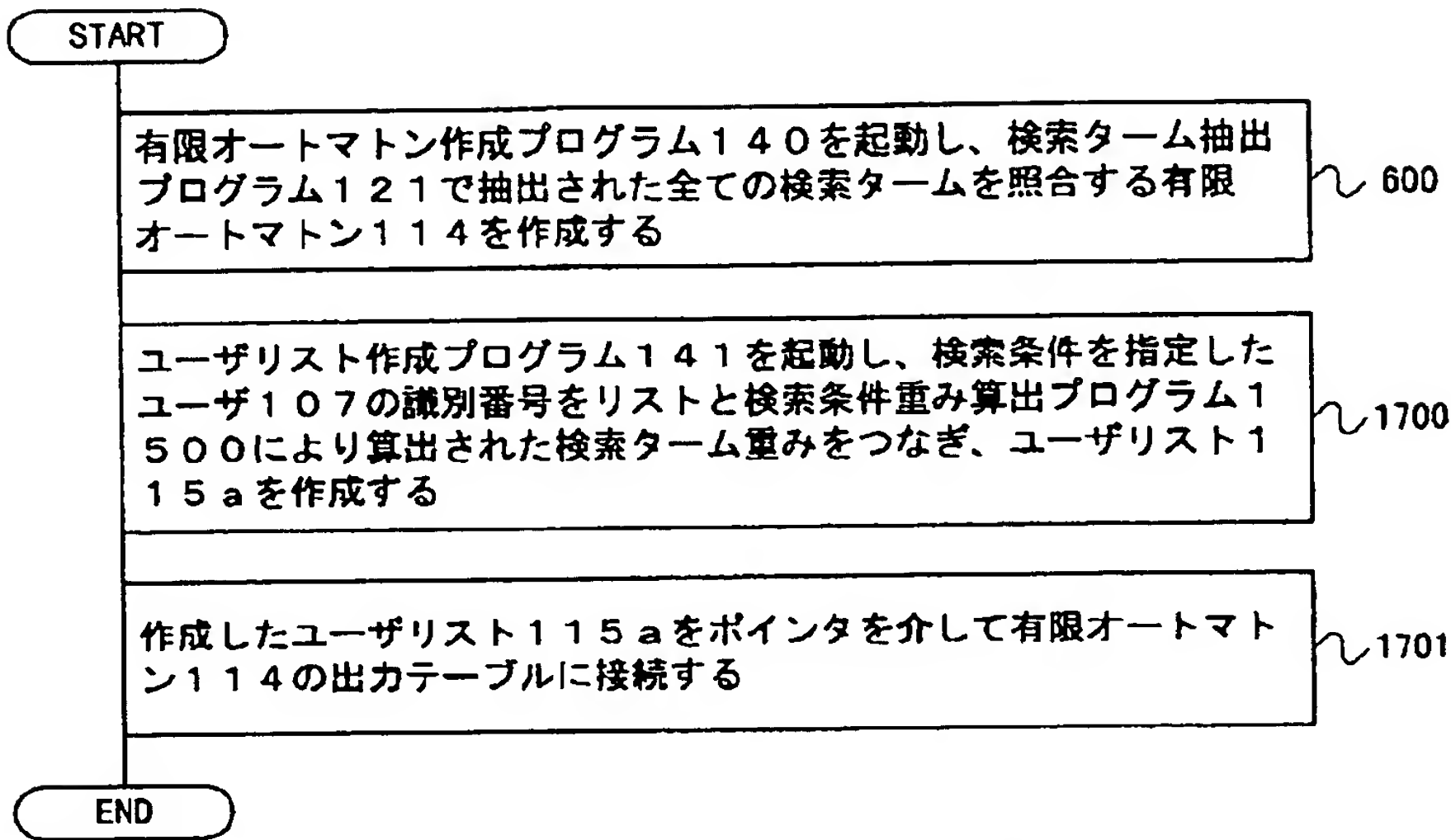
図 35



(追加配送テキスト数設定プログラム2900の処理手順)

【図 17】

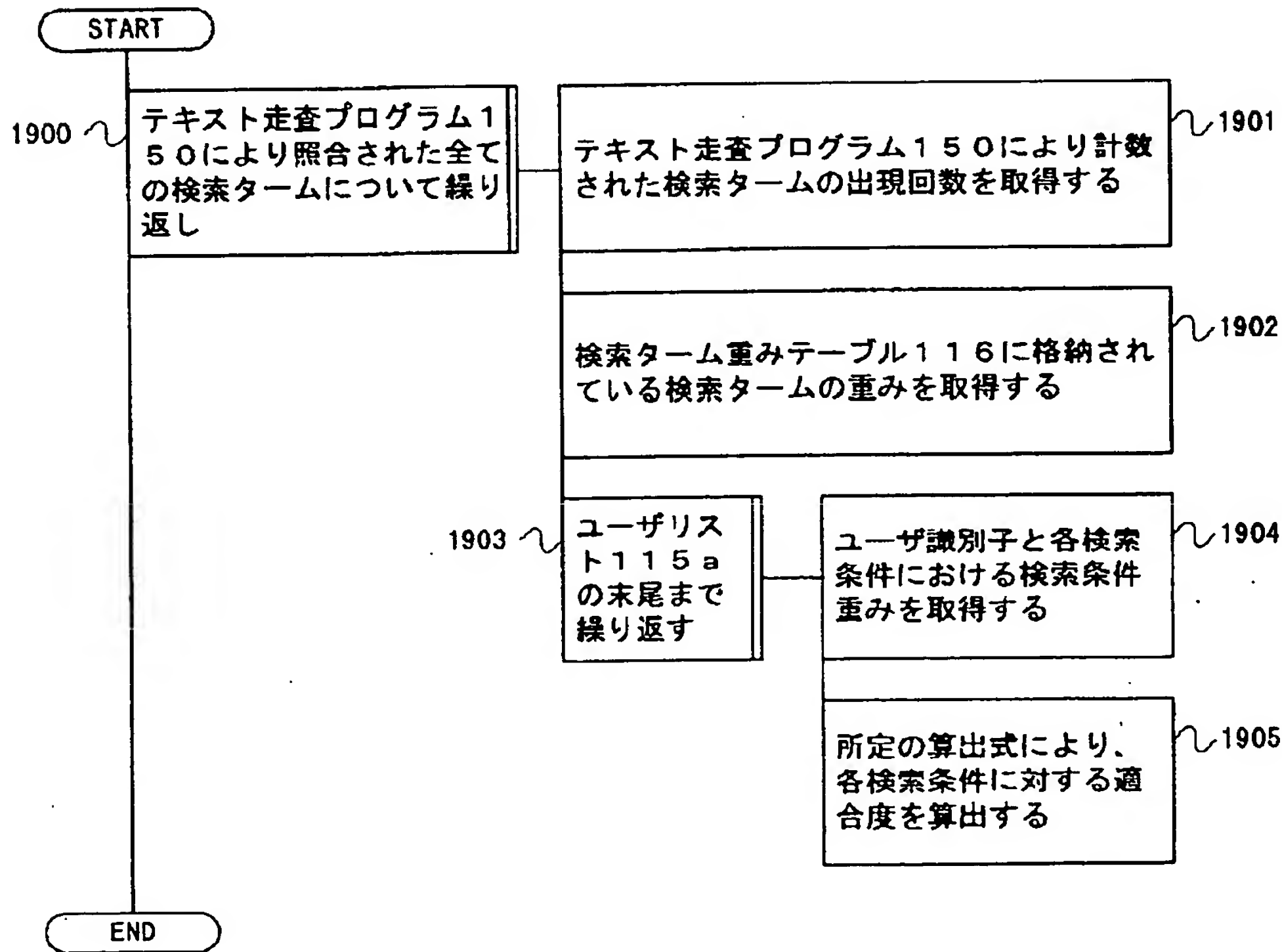
図 17



(本発明の第三の実施例における検索用オートマトン作成プログラム123aの処理手順)

【図 19】

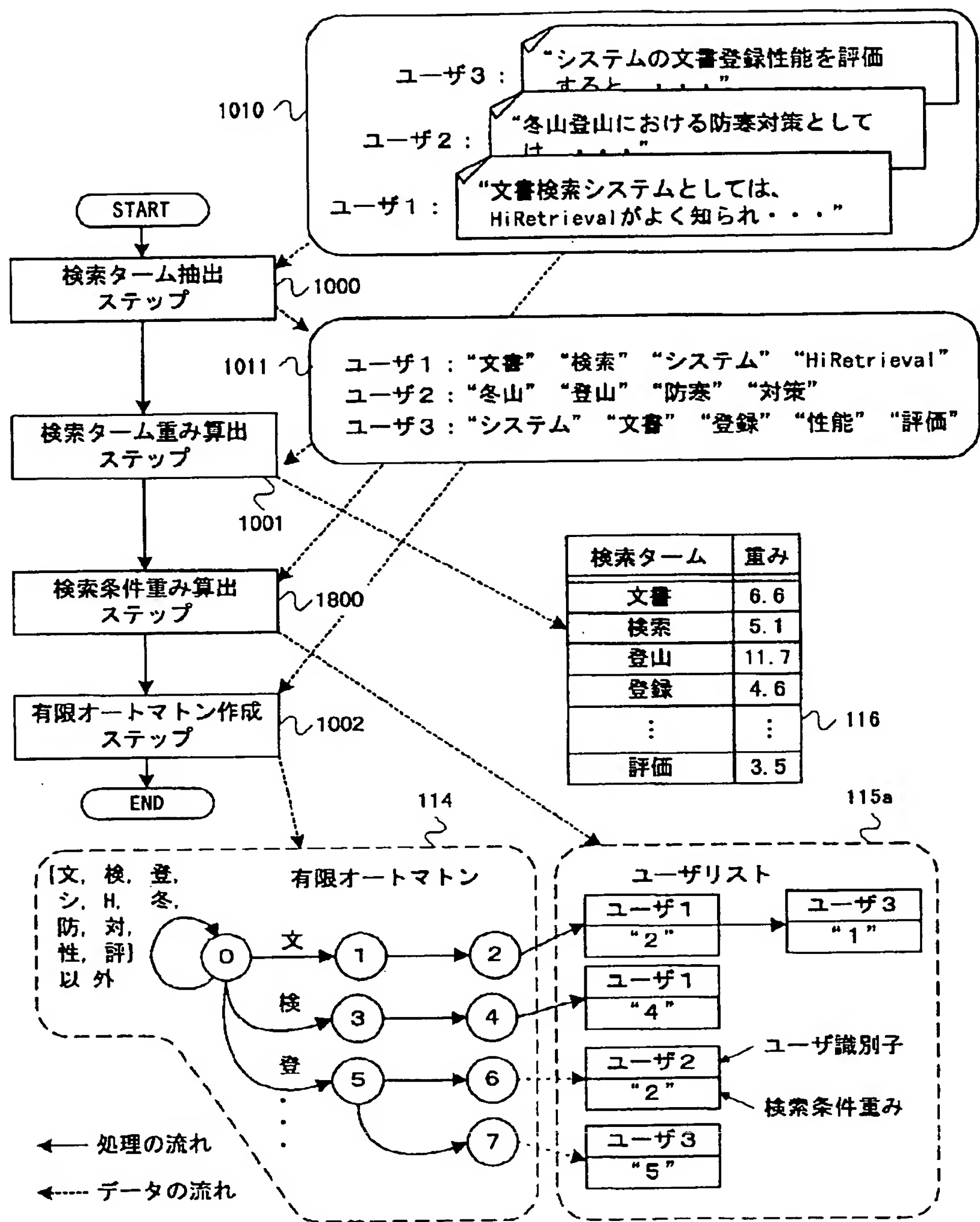
図 19



(本発明の第三の実施例における適合度算出プログラム151aの処理手順)

【図 18】

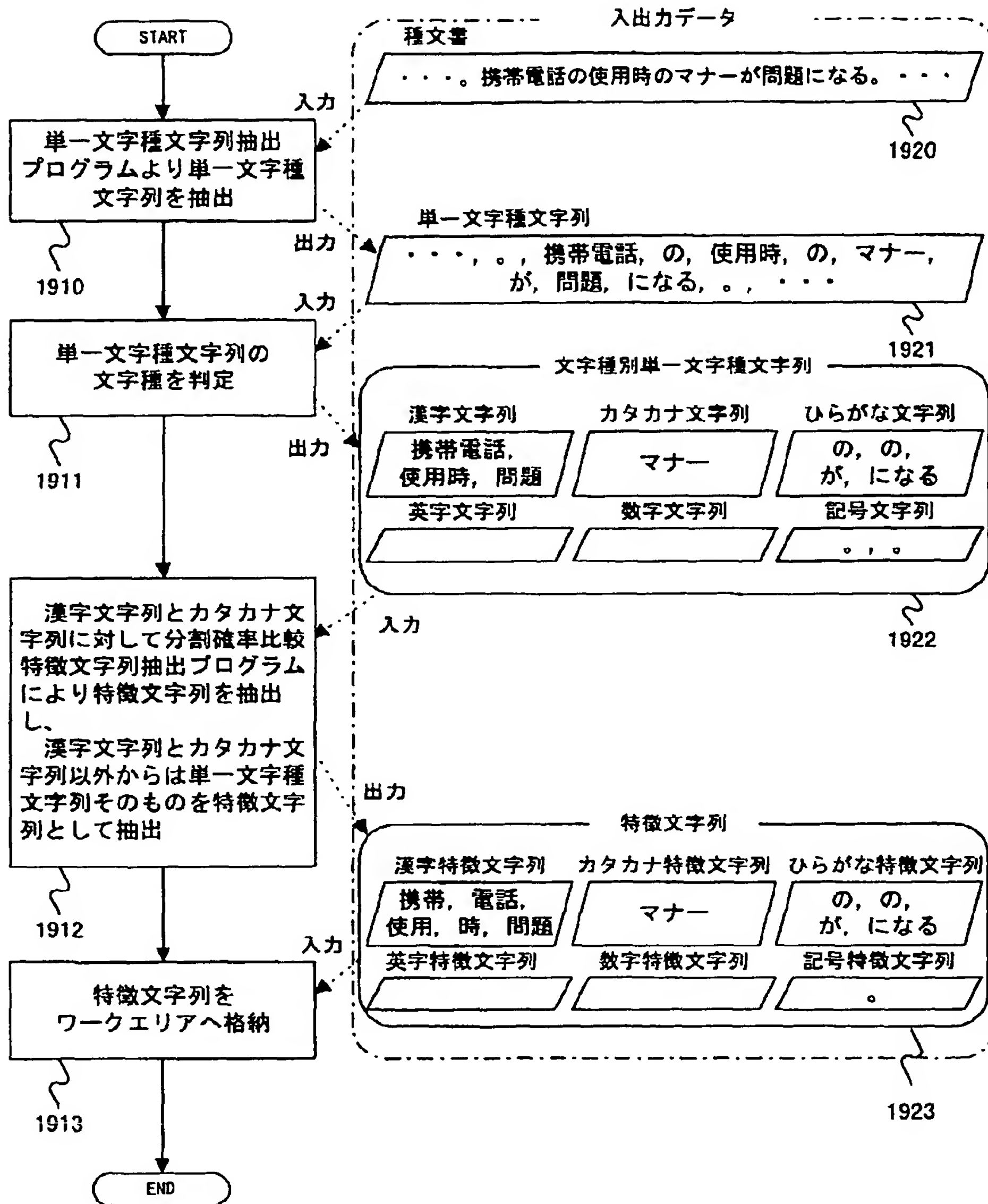
図 18



(検索条件登録制御プログラム111aの具体的な処理の流れを説明する図)

【図20】

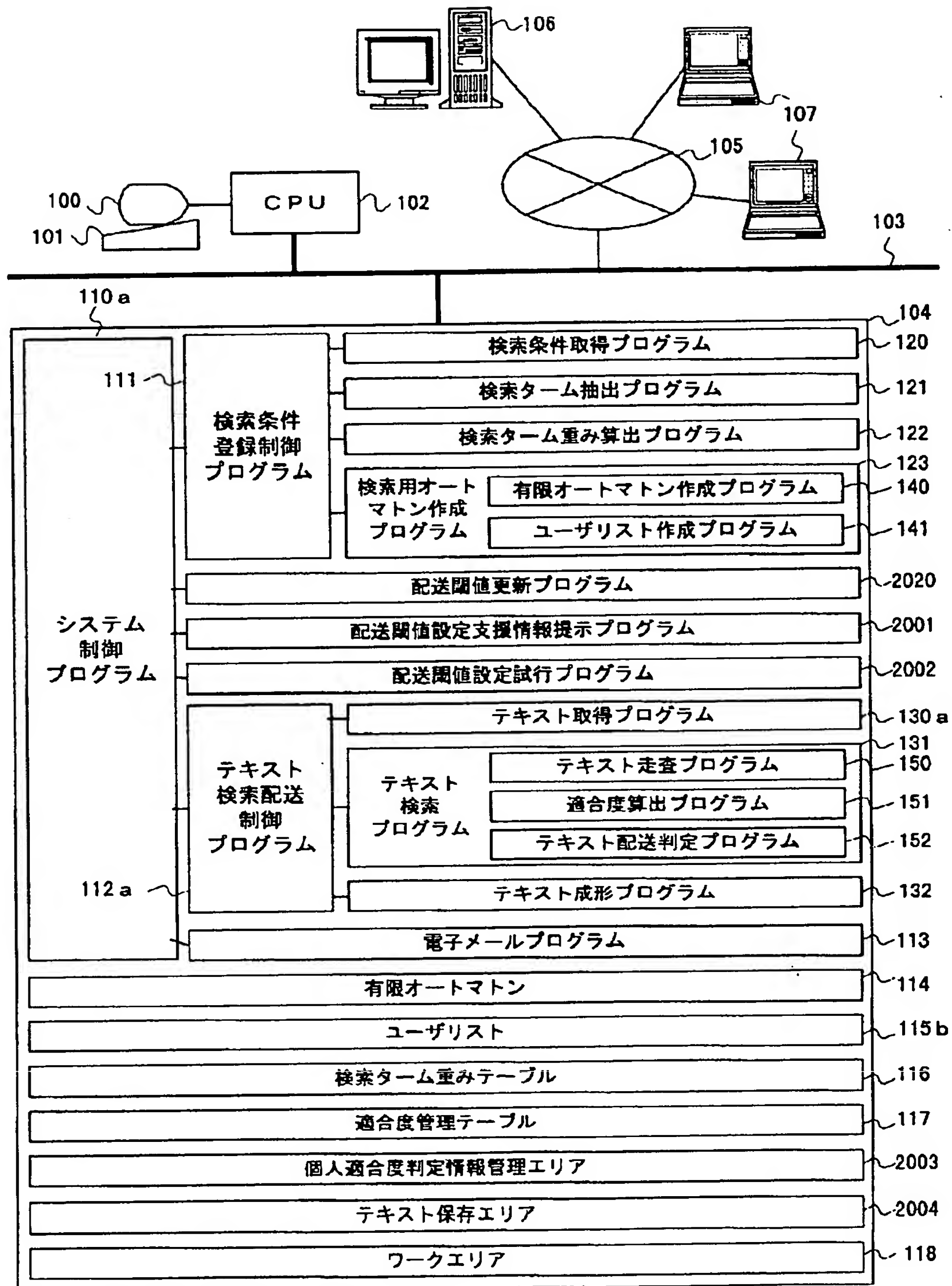
図20



(従来技術2における特徴文字列抽出処理の流れを説明する図)

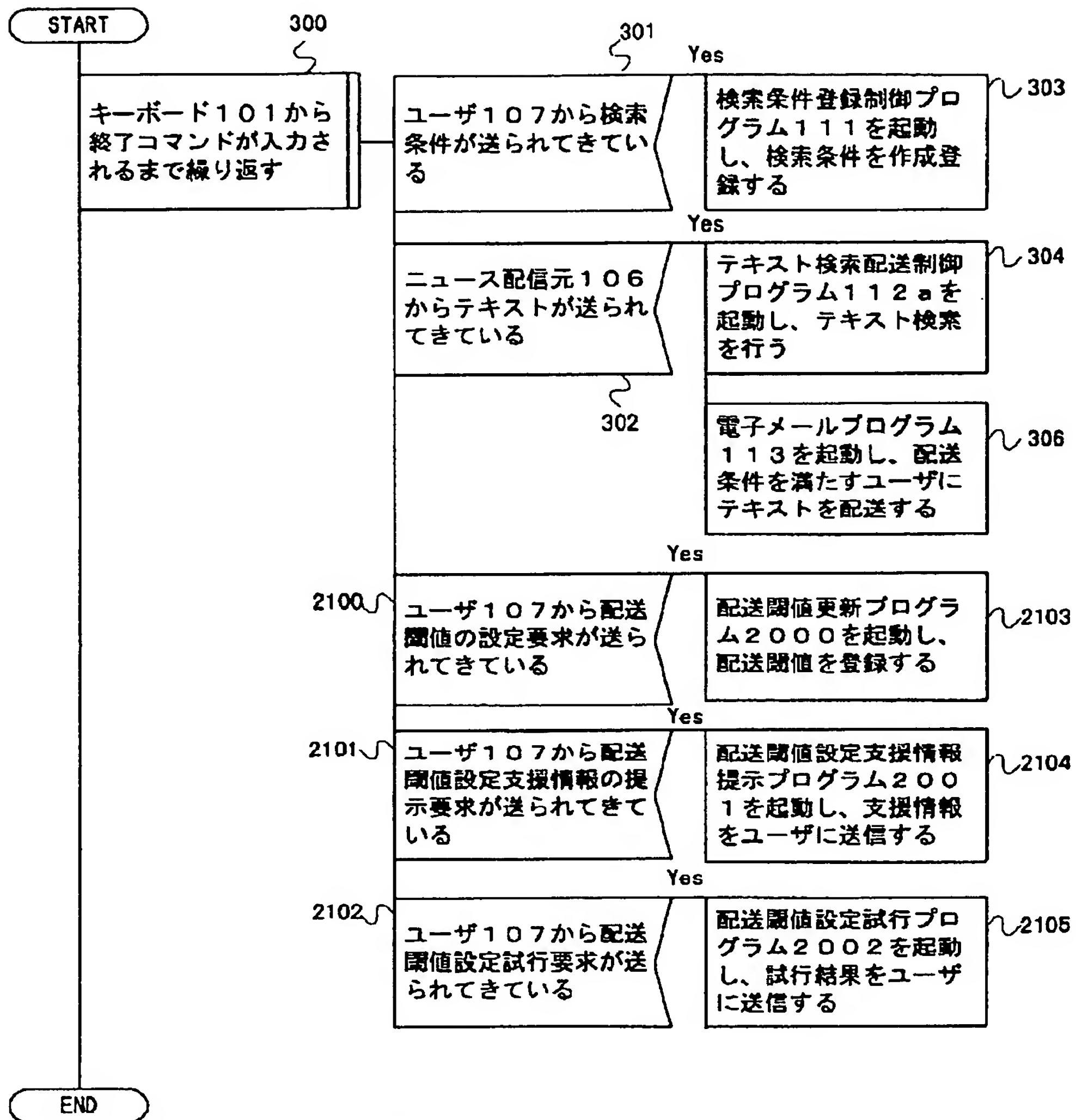
【図21】

図21



【図22】

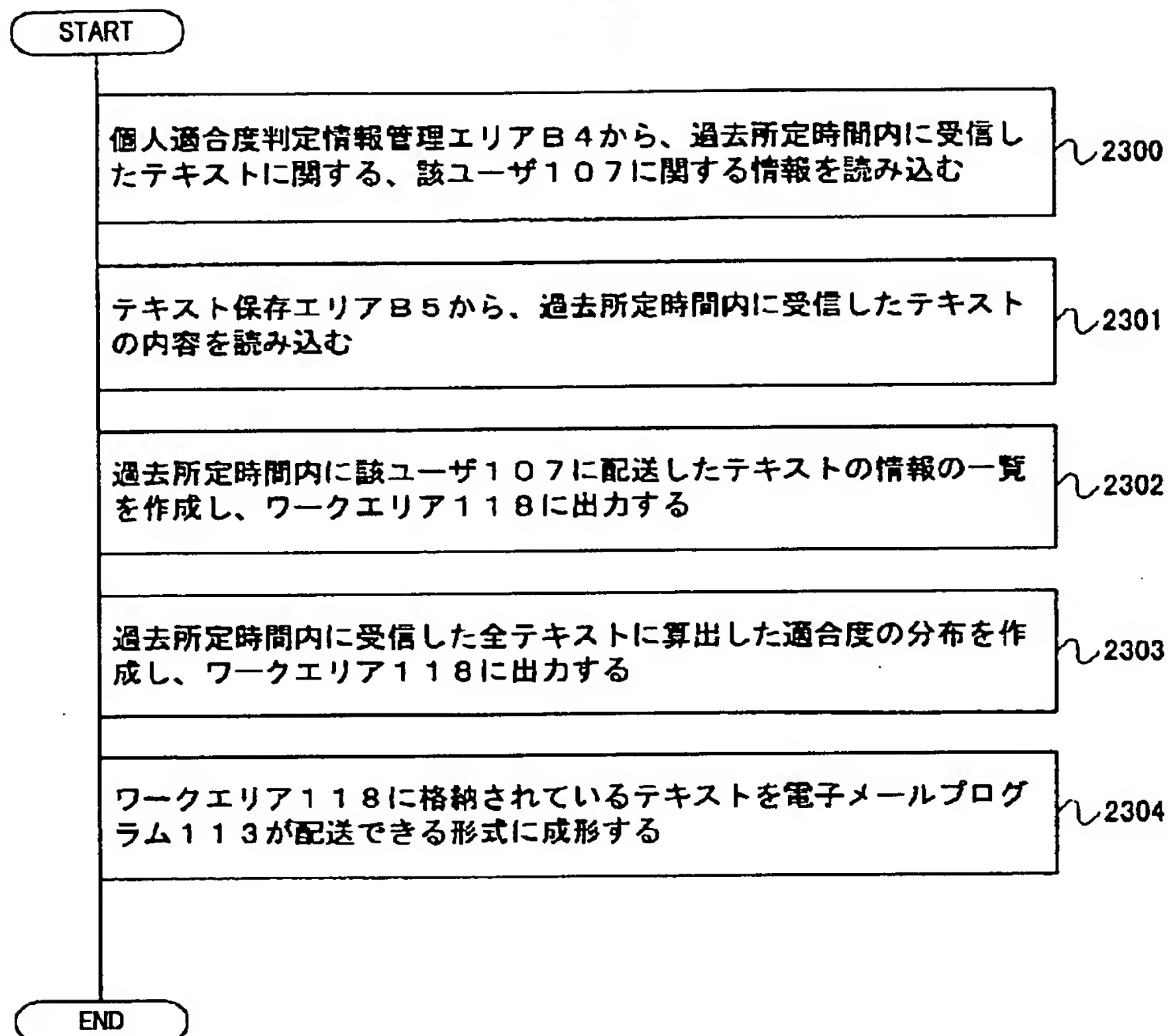
図22



(システム制御プログラム110aの処理手順)

【図 24】

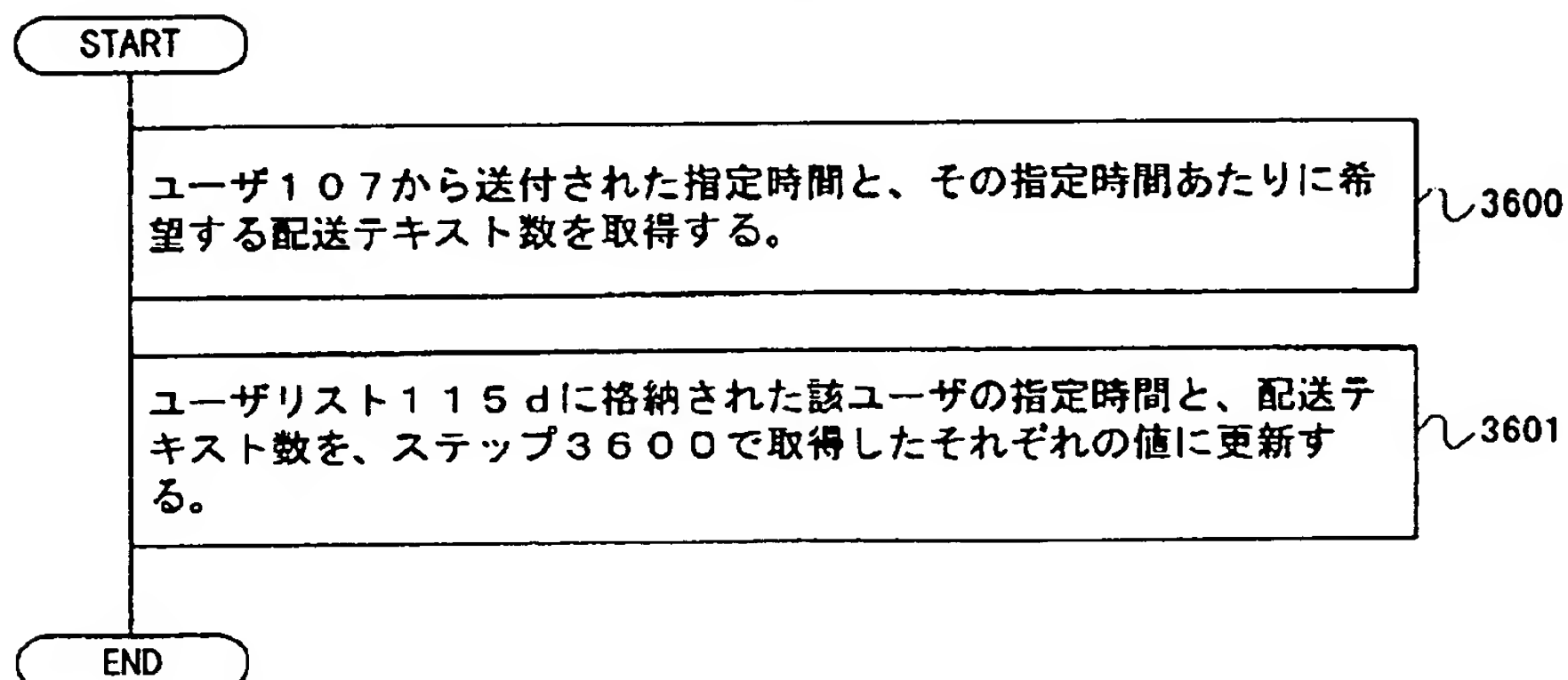
図 24



(配送閾値設定支援情報提示プログラム2001の処理手順)

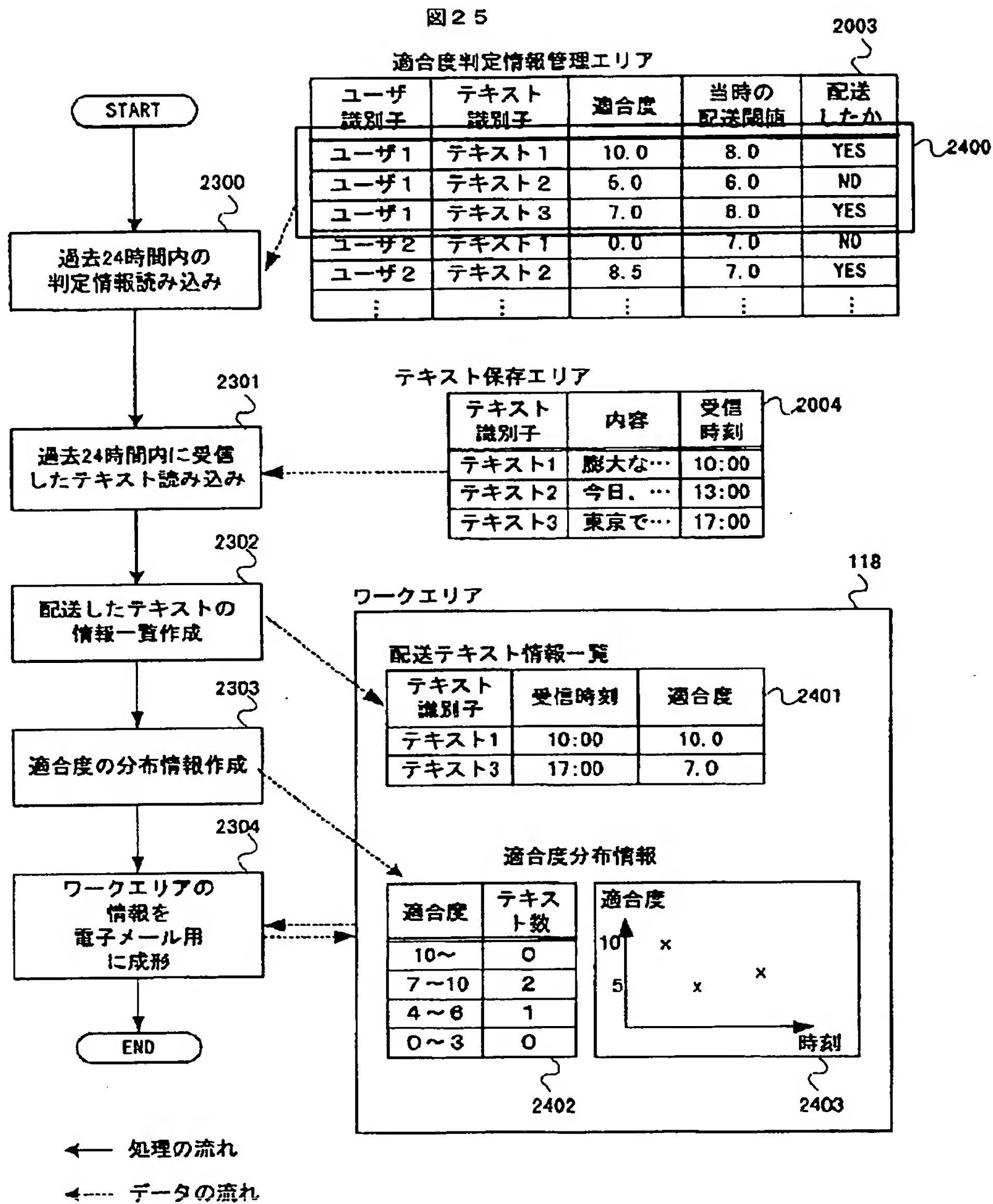
【図 40】

図 40



(希望配送数設定プログラム3400の処理手順)

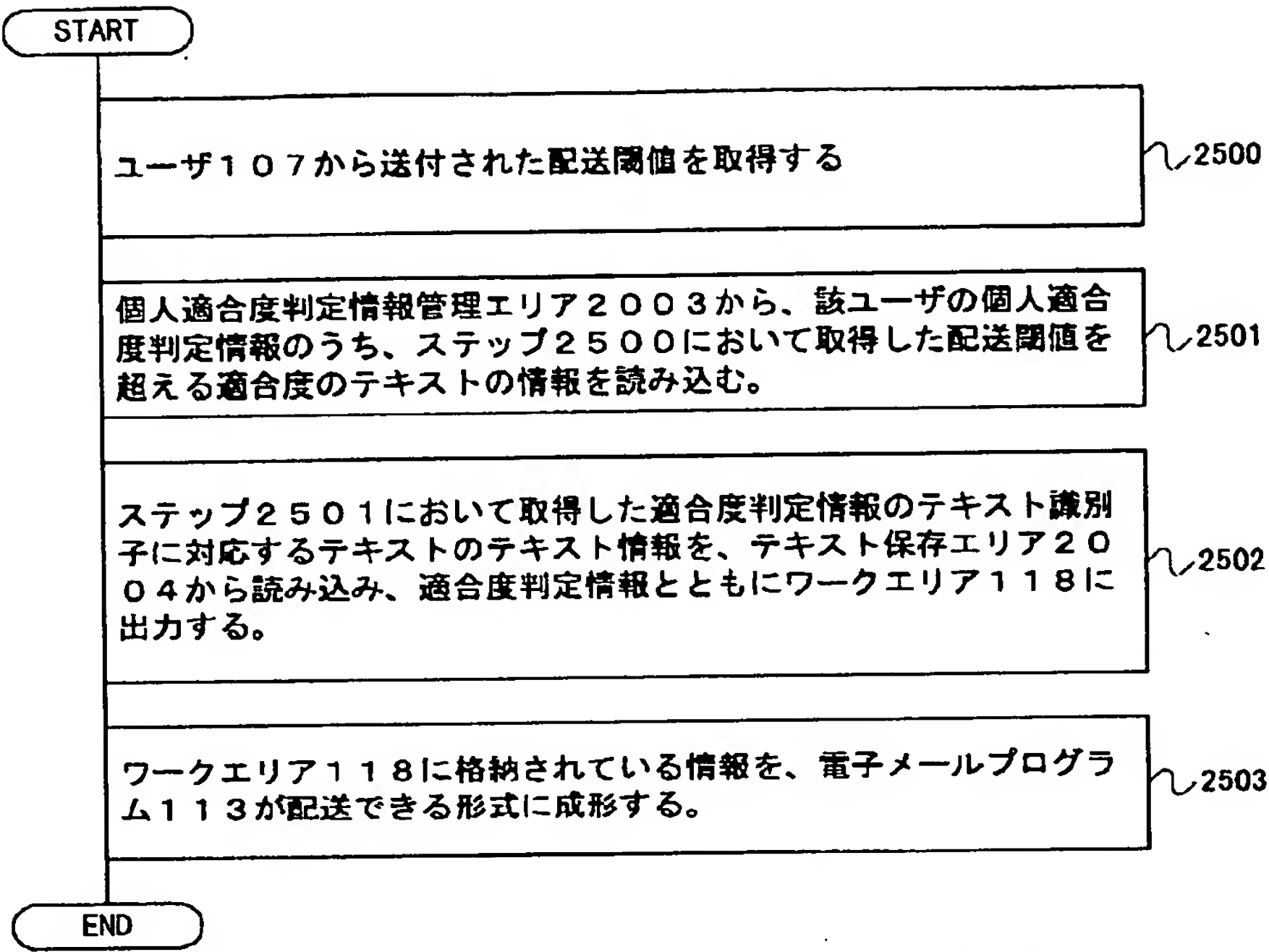
【図25】



(配送閾値設定支援情報提示プログラム2001の具体的な処理の流れを説明する図)

【図 28】

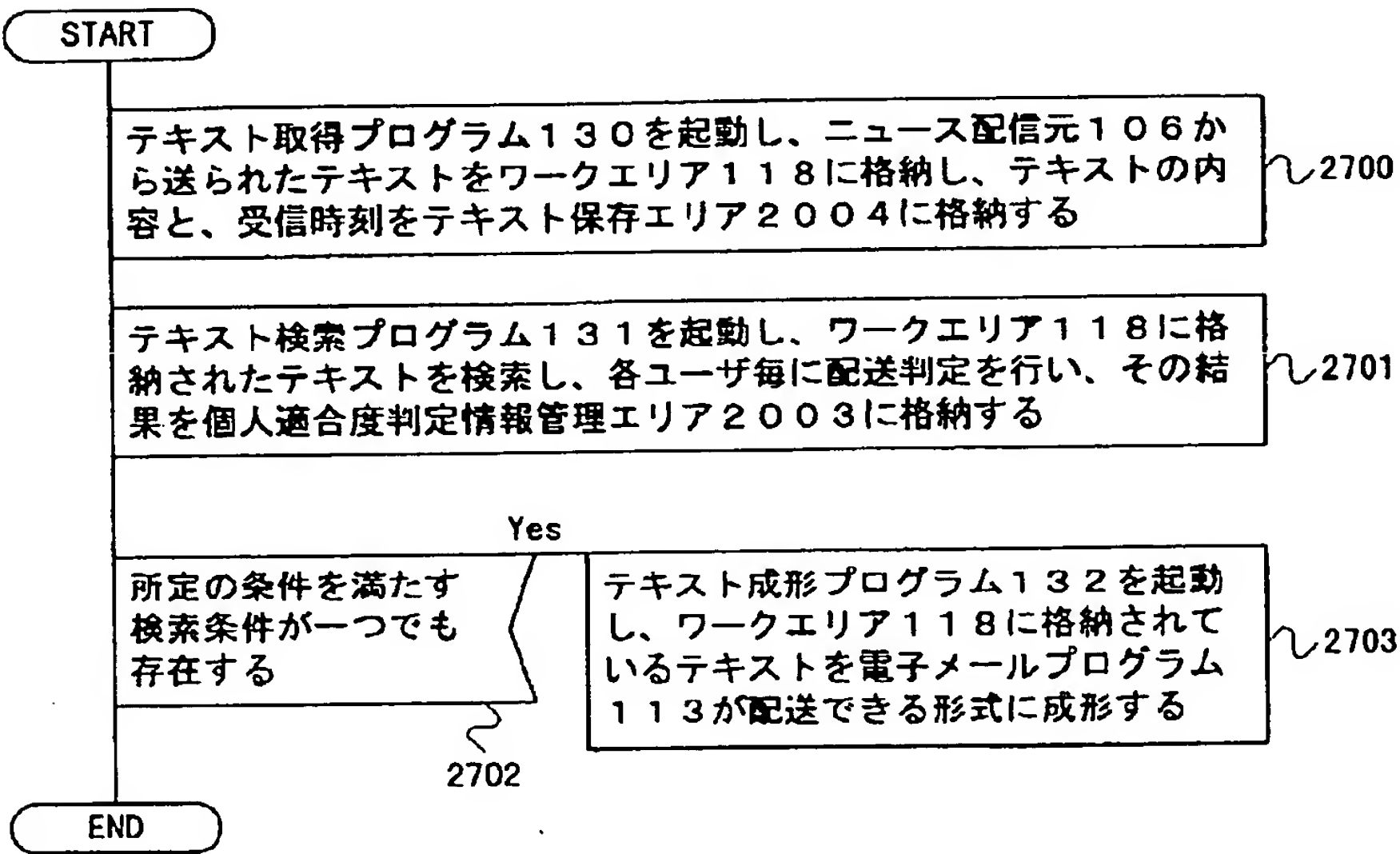
図 28



(配送閾値設定試行プログラム2002の処理手順)

【図 31】

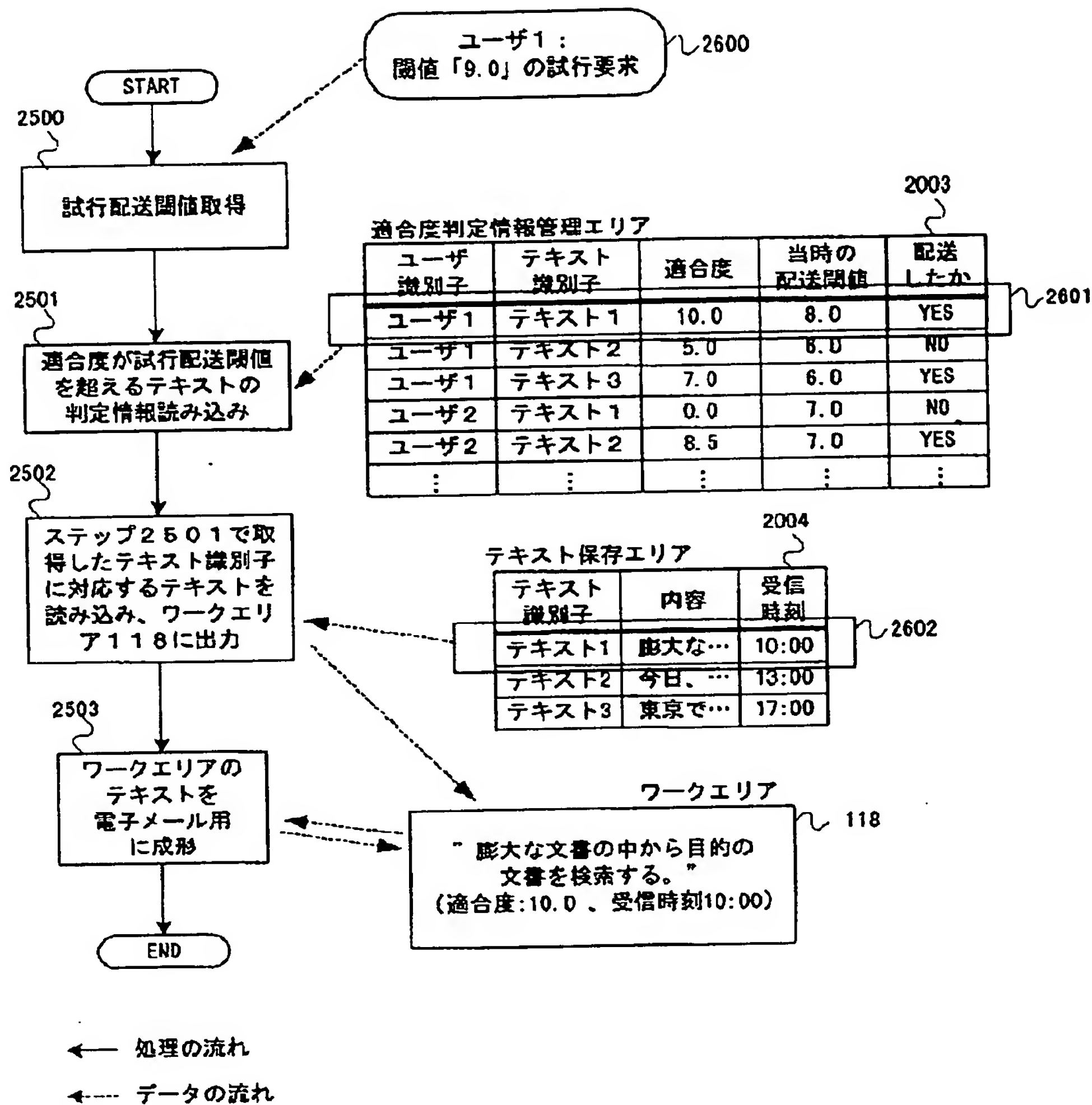
図 31



(テキスト検索配送制御プログラム112aの処理手順)

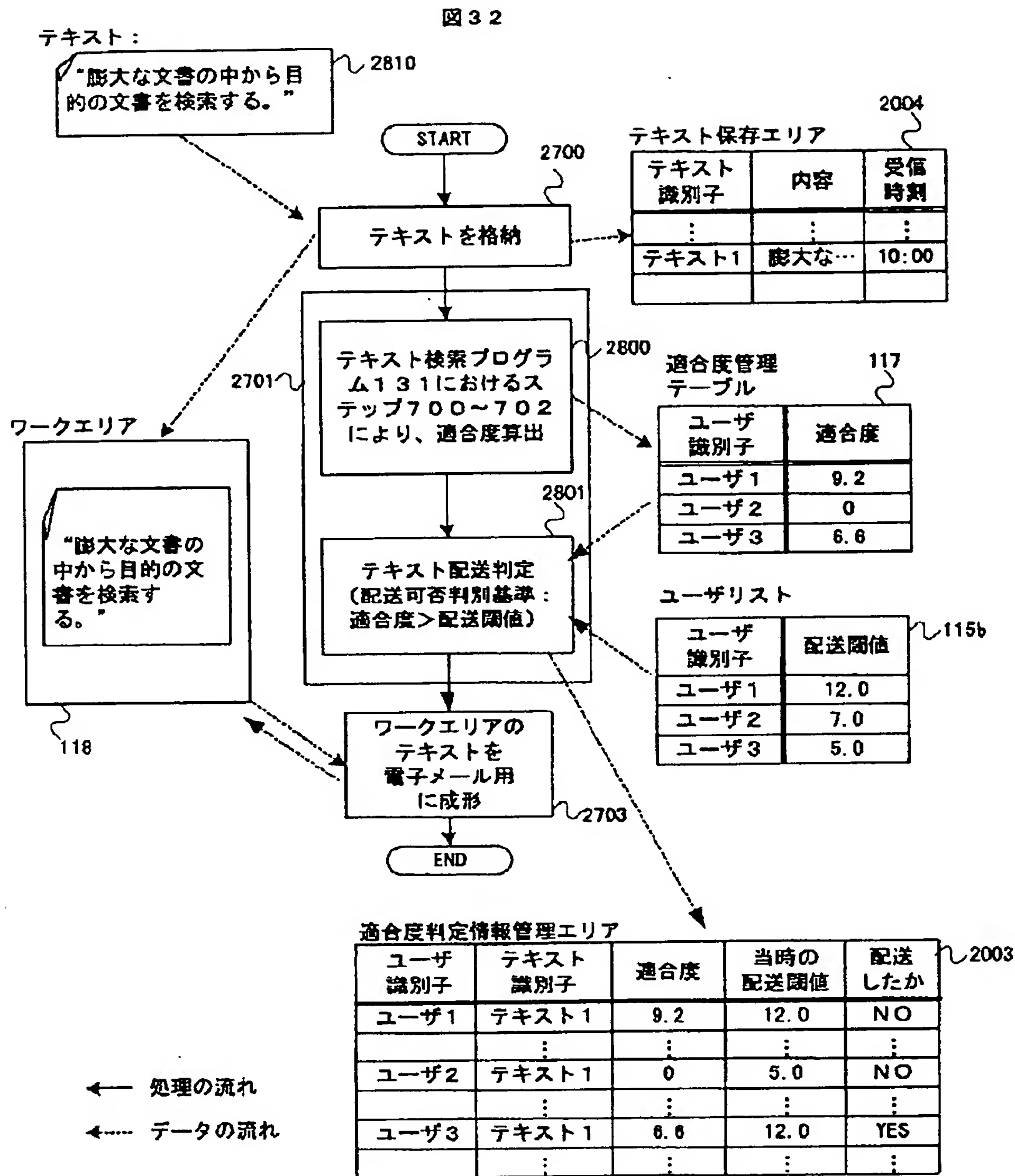
【図29】

図29



(配送閾値設定試行プログラム2002の具体的な処理の流れを説明する図)

【図32】



(テキスト検索配送プログラム112aの具体的な処理の流れを説明する図)

【図33】

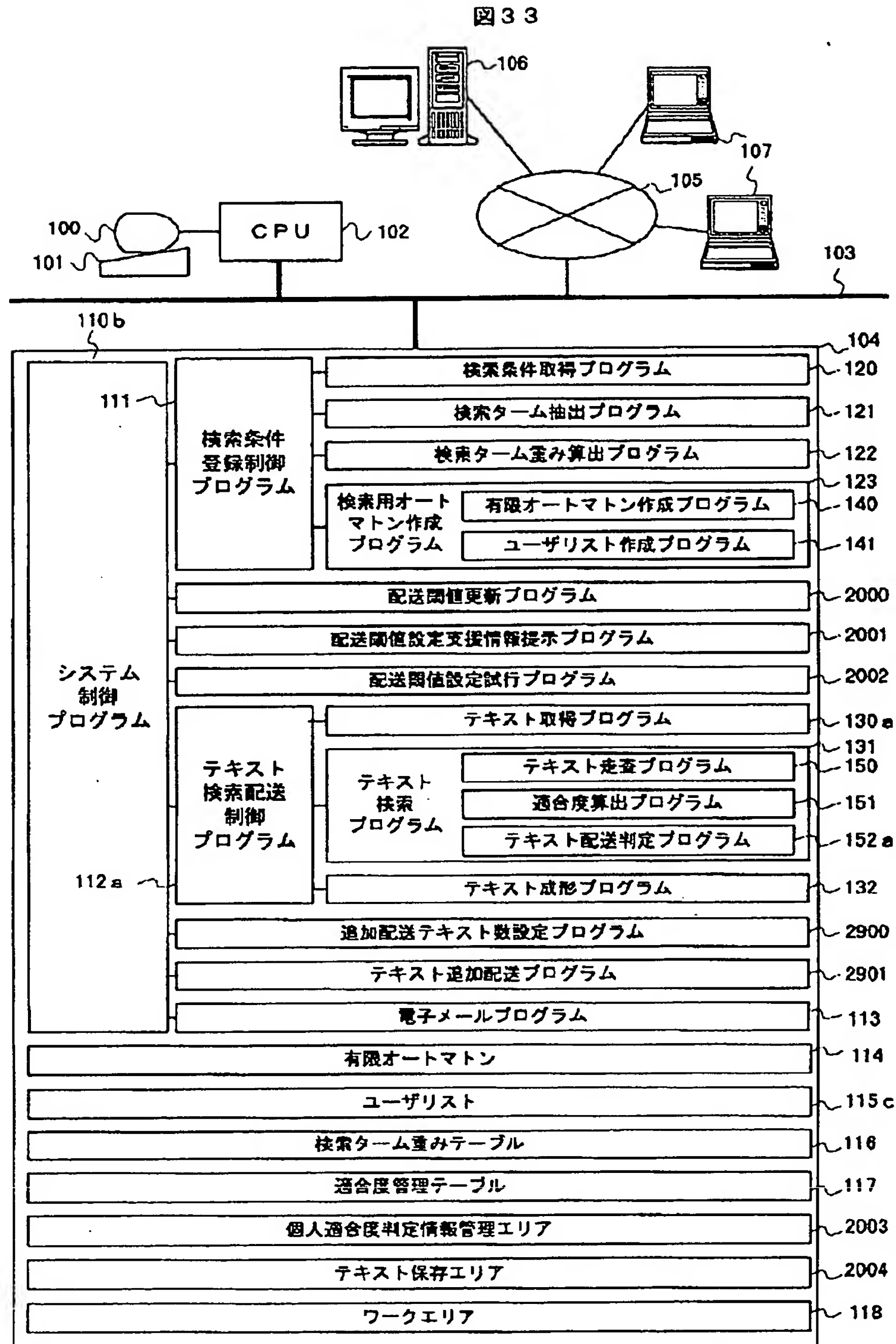
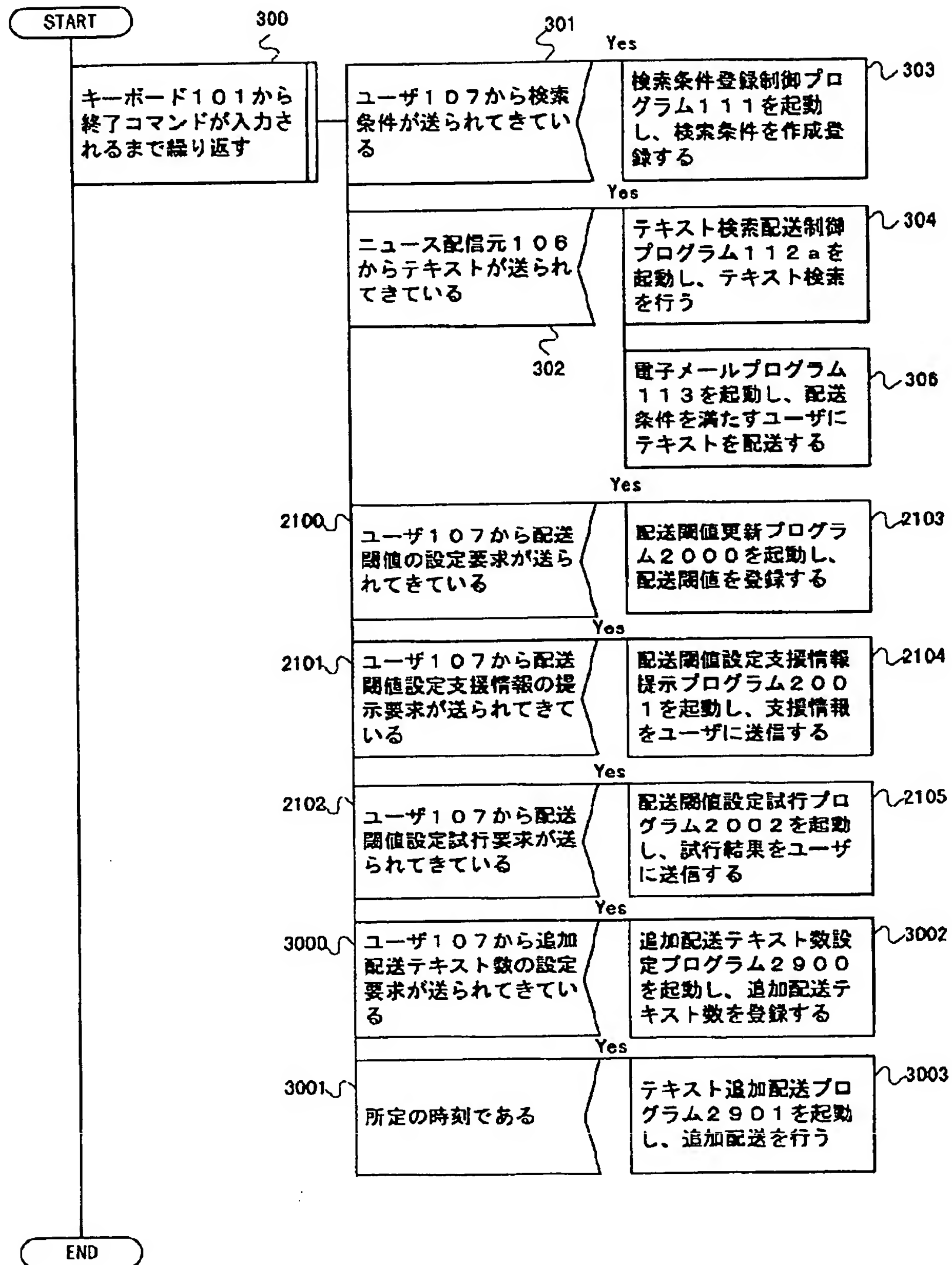
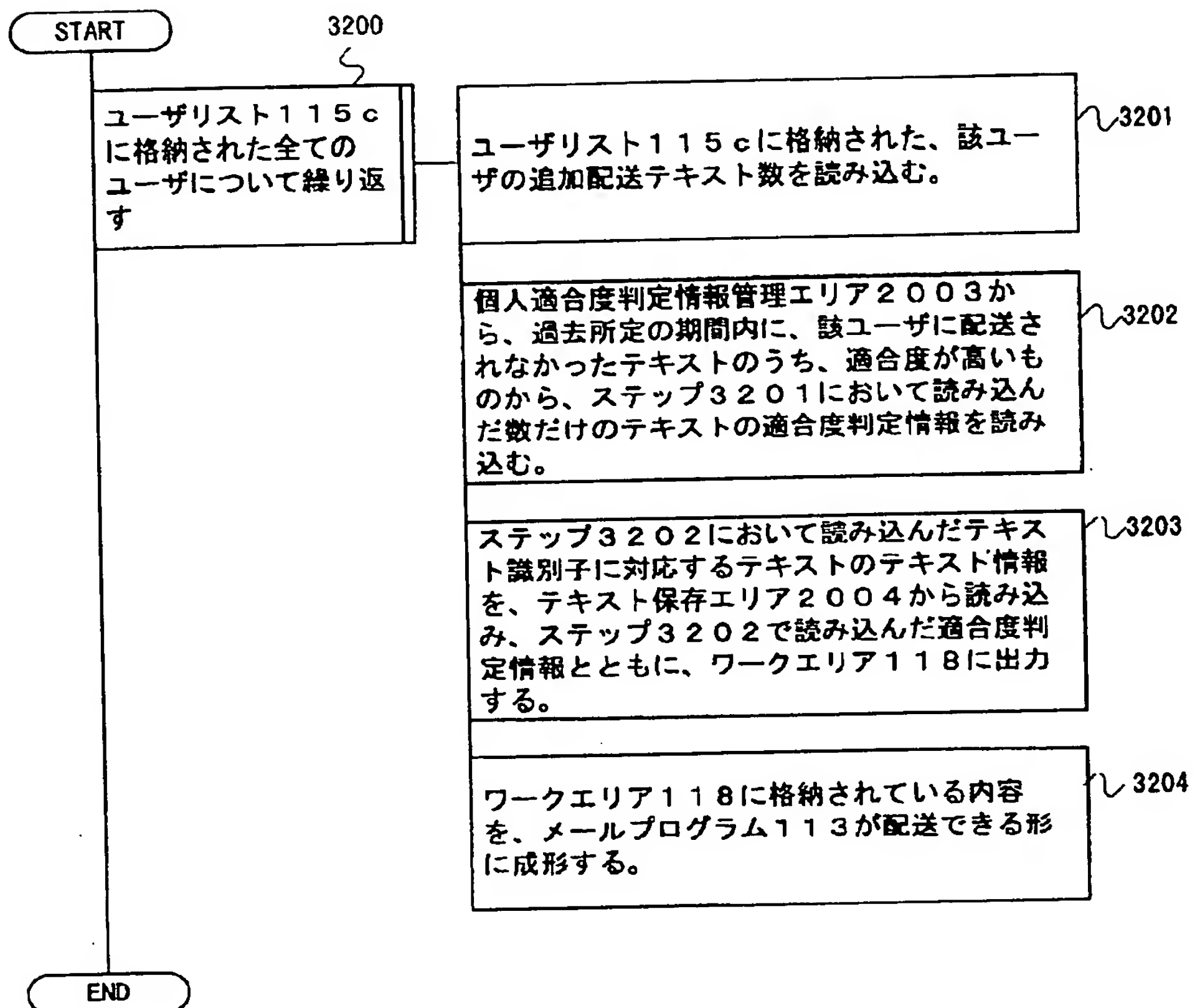


图 3 4



【図 36】

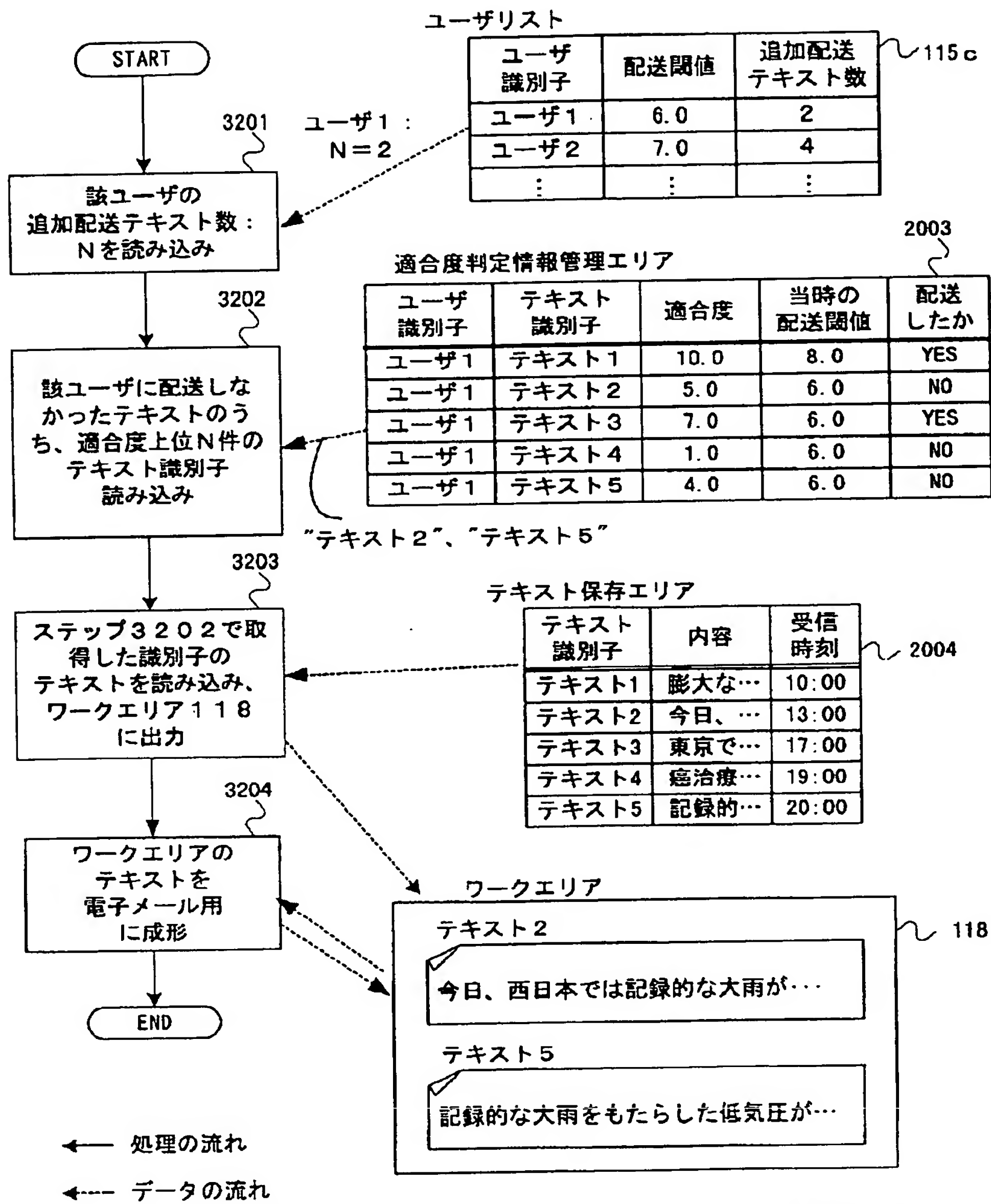
図 36



(テキスト追加配送プログラム2901の処理手順)

【図 3 7】

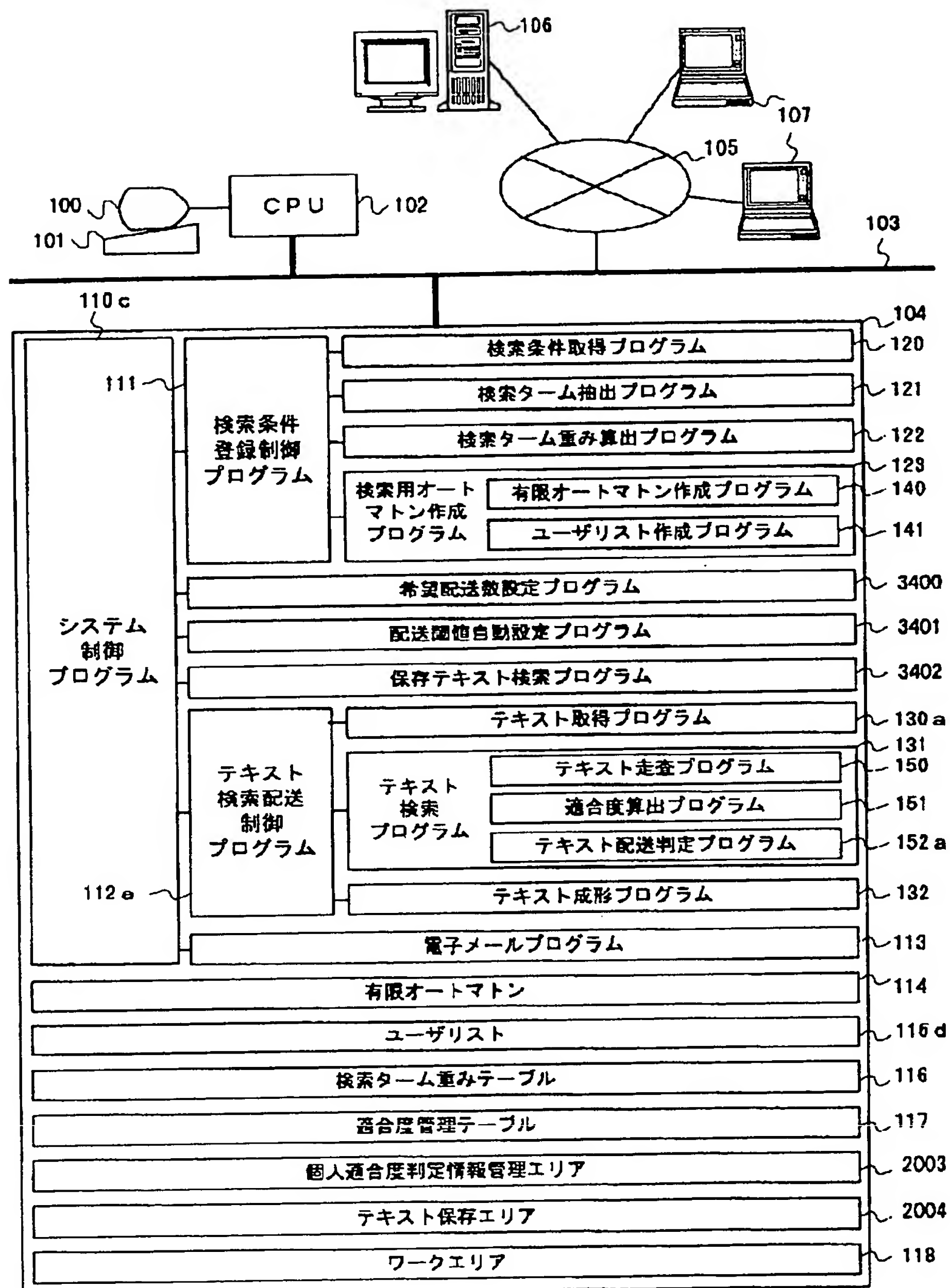
図 3 7



(テキスト追加配送プログラム 2 9 0 1 の具体的な処理の流れを説明する図)

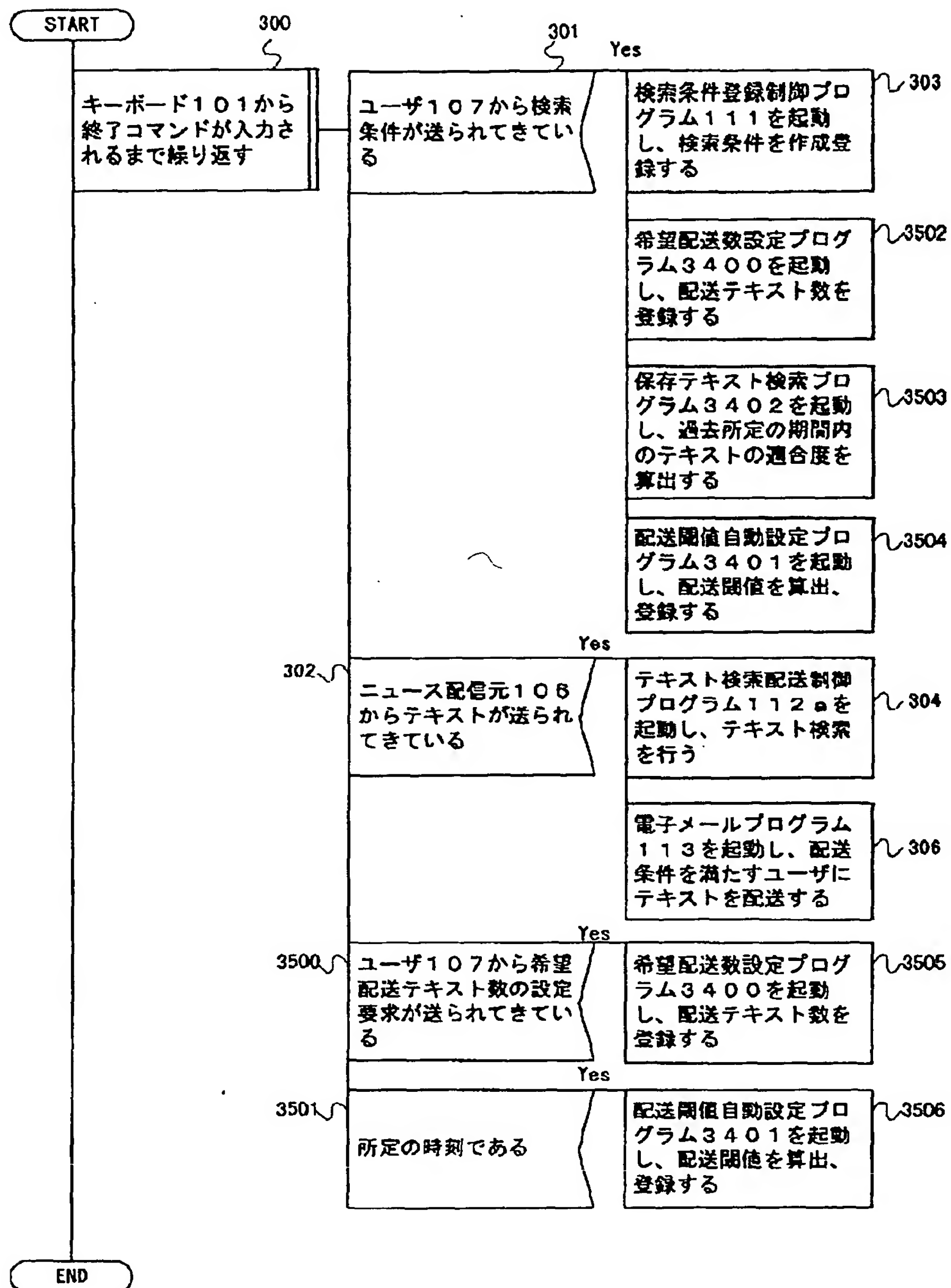
【図 38】

図 38



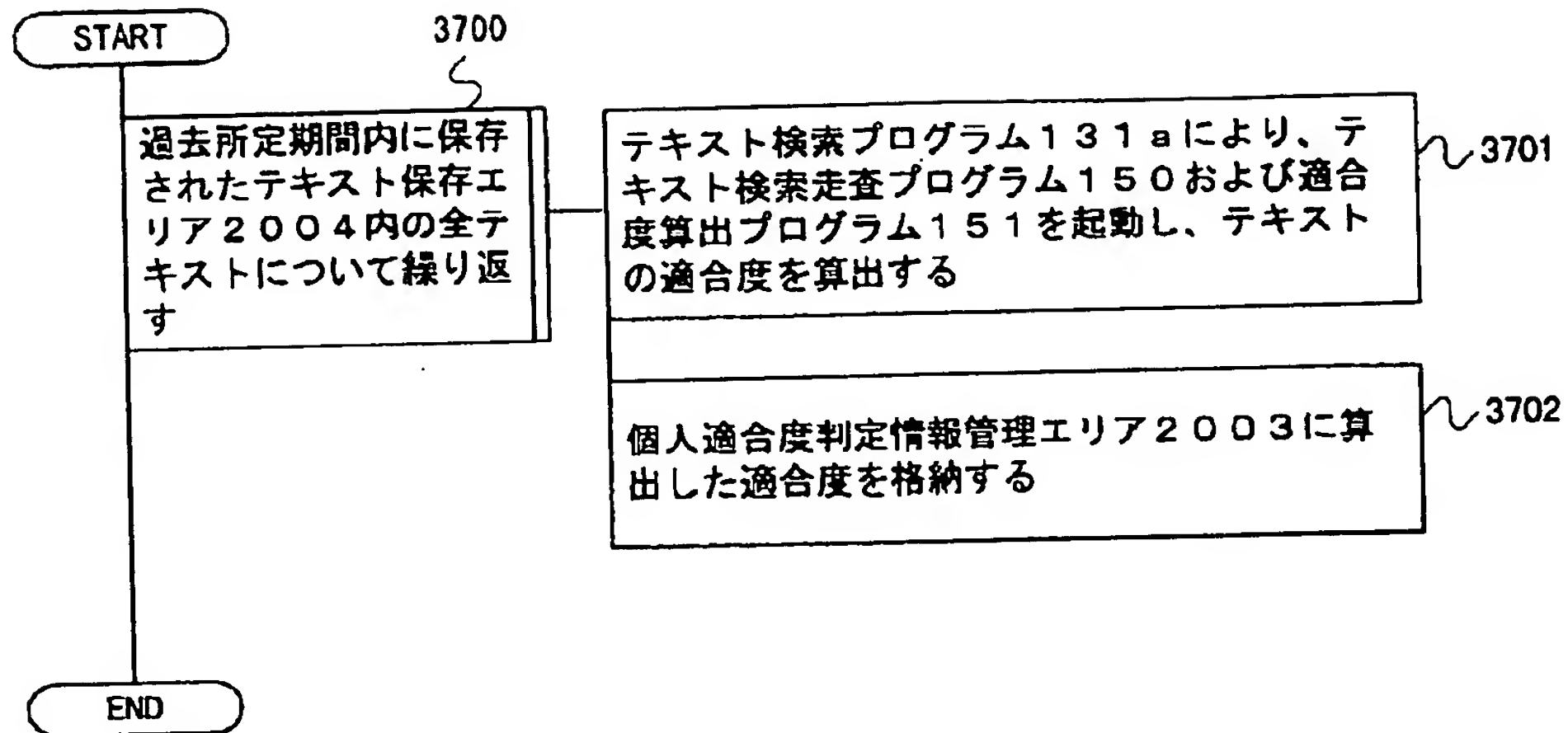
【図39】

図39



【図 41】

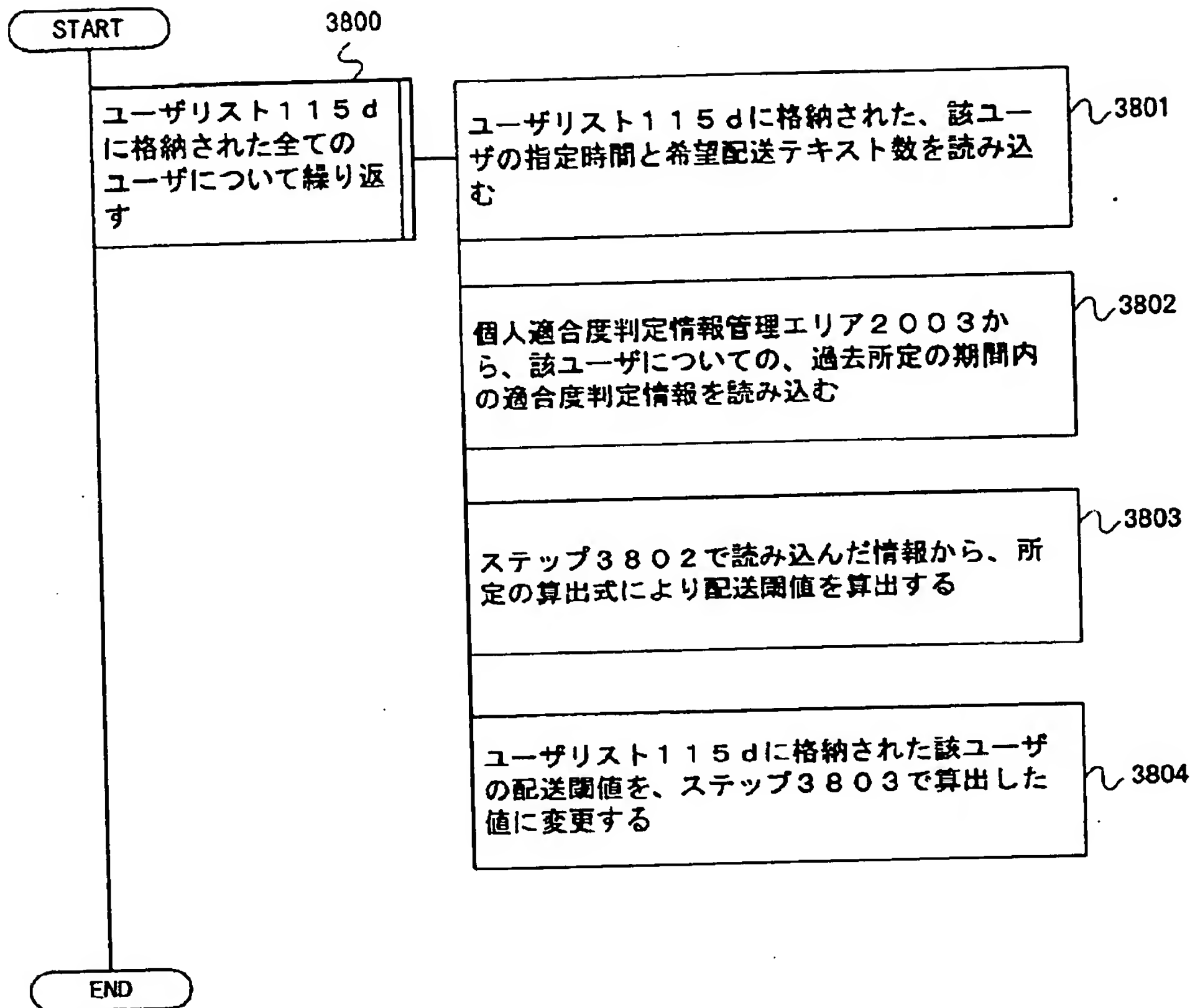
図 41



(保存テキスト検索プログラム3402の処理手順)

【図42】

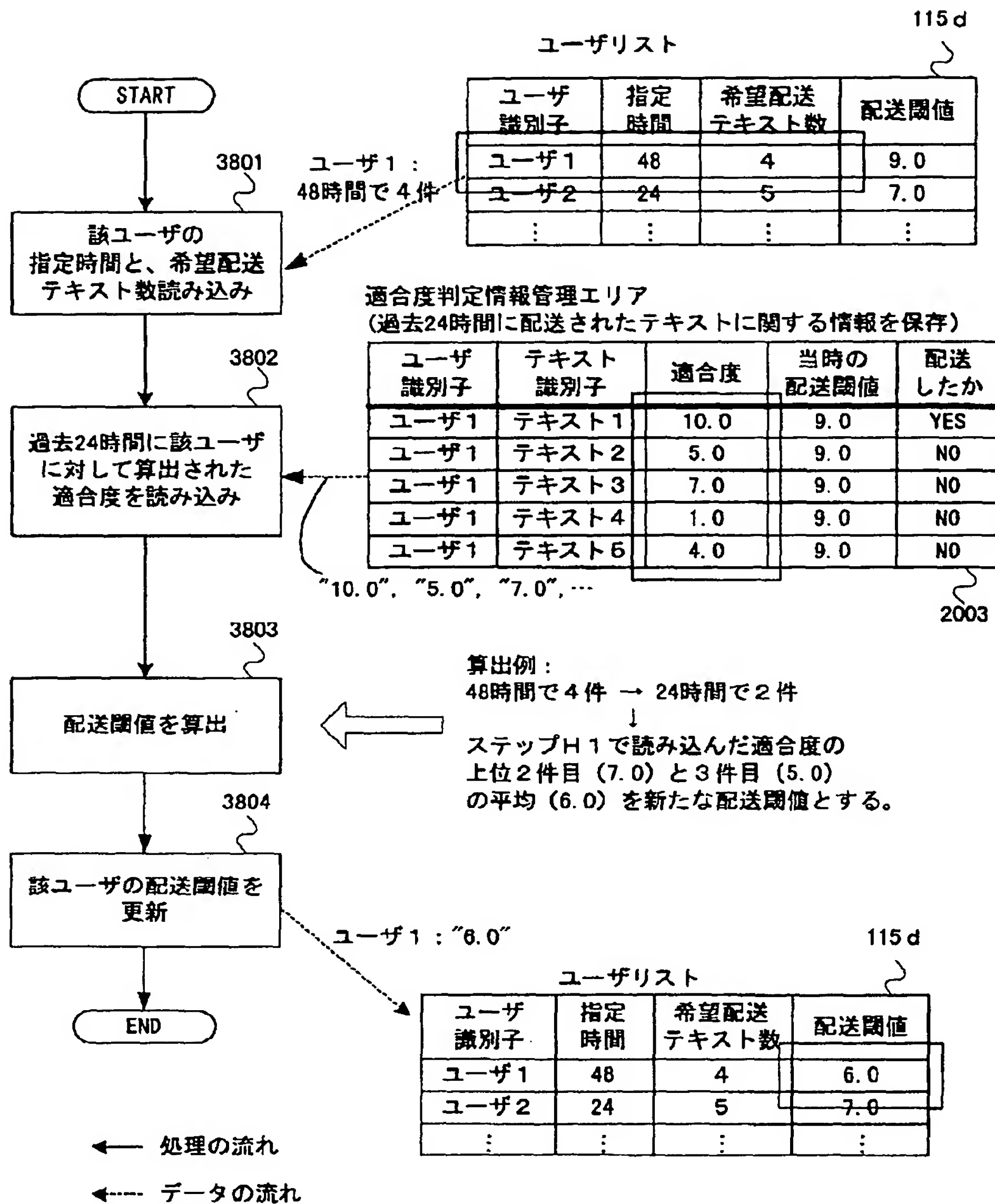
図42



(配送閾値自動設定プログラム3401の処理手順)

【図43】

図43



(配送閾値自動設定プログラム3401の具体的な処理の流れを説明する図)

フロントページの続き

(72)発明者 多田 勝己
 神奈川県川崎市幸区鹿島田890番地 株式
 会社日立製作所システム開発本部内

(72)発明者 岡本 卓哉
 神奈川県川崎市幸区鹿島田890番地 株式
 会社日立製作所システム開発本部内

(72) 発明者 菅谷 奈津子
神奈川県川崎市幸区鹿島田890番地 株式
会社日立製作所システム開発本部内

(72) 発明者 後地 陽介
神奈川県横浜市戸塚区戸塚町5030番地 株
式会社日立製作所ソフトウェア事業部内
F ターム (参考) 5B075 ND03 NK32 PP24 PQ05 PQ74
PR06 QM08 QS01